PCT/EP 99/04079

EP99/04079

Europäisches Patentamt | European Patent Office | Office européen des brevets

5

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

99107554.0

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

Arlette Fiedler

A. Fiedler

DEN HAAG, DEN
THE HAGUE, 27/07/99
LA HAYE, LE

Europäisches Patentamt | European Patent Office | Office européen des brevets

Blatt 2 d r B sch inigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 99107554.0
Demande n°:

Anmeldetag:
Date of filing: 15/04/99
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Biosearch Italia S.p.A.
21040 Gerenzano (VA)
ITALY

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Methods for transferring the capability to produce a natural product into a suitable production host

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

METHODS FOR TRANSFERRING THE CAPABILITY TO PRODUCE A NATURAL PRODUCT INTO A SUITABLE PRODUCTION HOST.

1. FIELD OF THE INVENTION

The present invention relates to a novel approach for drug discovery. More particularly, the invention relates to a system for improving the process of lead optimization and development of compounds, when these compounds are natural products produced by microorganisms belonging to the order *Actinomycetales* or chemical derivatives of these compounds. The invention relates to a system for transferring the capability to produce a natural product from a microorganism belonging to the order *Actinomycetales* into a defined host, where said natural product can be optimally produced and its biosynthetic pathway suitably modified.

2. BACKGROUND ART

Natural products are complex molecules with important uses in medicine. Examples include: antibacterial agents, such as erythromycin, teicoplanin, tetracycline; antitumor compounds, such as dauxorubicin; antihelmintic compounds, such as avermectin; immunosuppressive agents, such as cyclosporin and FK506; antifungal compounds, such as amphotericin and nystatin; etc. Natural products are

produced as secondary metabolites by a wide range of living organisms. Although many secondary metabolites have been identified, there remains the need to obtain novel structures with new activities or enhanced properties. Current methods of obtaining such molecules include screening of natural isolates and chemical modification of existing ones. Random screening of natural products from disparate sources has resulted in the discovery of many important drugs and is still employed for seeking for novel activities. This process, which consists in exposing a miniaturized biological system to tens or hundreds of thousands of different compounds, in order to find those few that exhibit a desired property, is designated high throughput screening, or HTS.

One of the used sources widely in HTS is a collection of natural products produced by small-scale fermentation of newly isolated microorganisms. A natural product may have one or more potential therapeutic properties, including but not limited to antibacterial, antifungal, antiviral, antitumor, immunomodulating or other pharmacological properties. Natural products have long constituted a source of interesting, structurally original and "imaginative" molecules endowed with potent biological activities. In addition, recent observations indicate that only a small fraction of the microbial flora present in environmental samples, ranging from 0.01 to 1% according to the estimates, is related to known species. Microorganisms belonging to the order

Actinomycetales represent thus far the group of producers unsurpassed for chemical and biological diversity. However, more than 15,000 natural products produced by microorganisms have been described, and the chances of finding new structures are relatively small, unless efforts are directed towards those classes of microorganisms that have been little exploited in the past. Poorly characterized actinomycete genera can thus constitute a useful source of novel structures. With proper methodologies, unusual genera can be isolated from environmental samples and some of these isolates will produce interesting activities. These could either represent completely new entities, or known molecules acting on a novel target or in a previously unreported way. Many of these products will have original structures and potent biological activities. However, newly discovered secondary metabolites will be produced for the most part by microorganisms which have been isolated for the characteristic of being unusual and selected for their ability to produce a given bioactivity. Consequently, little will be known about the best conditions for growth, productivity and storage. Often the microorganism does not produce a single bioactive compound, and other, unrelated activities must be completely removed for a meaningful evaluation of the properties of the lead compound. Furthermore, rarely is a secondary metabolite produced as a single, bioactive molecule, but is often present as a "complex" of several, closely related compounds, only some of which may possess the desired biological

or chemical properties. Therefore, physiological conditions, such as nutrient and cofactor supply, that allow obtaining a "controlled" complex need to be established empirically by a trial and error approach. Finally, the natural product may need structural modification, and this can be achieved only by chemical means. In essence, the scarce knowledge available on the physiology and genetics of the producing strain will severely hamper the lead optimization and development processes.

Chemical modification of preexisting natural products has been successfully employed to generate derivatives of natural products, but it still suffers from practical limitations to the type of compounds obtainable. Many natural products are often structurally complex molecules, with relatively large molecular weights. Due to their structural complexity, total synthesis of natural products is often prohibitive for the number of necessary steps and the overall yield; furthermore, selective modification of a natural product can often be efficiently performed only on limited portions of the molecule. This difficulty of generating structural derivatives by conventional medicinal chemistry slows down the process of lead optimization and supply. Microorganisms employ intricate biosynthetic machineries to make natural products: for example, synthesis of the macrolide antibiotic erythromycin, a secondary metabolite in the medium-range structural complexity, requires the participation of over 40 different enzymatic activities

(Katz and Donadio, 1995, Macrolides, in Genetics and Biochemistry of Antibiotic Production, Vining and Stuttard eds., Butterworth-Heinemann, Boston CT, p. 385-420). Biosynthetic pathways can often be redirected through manipulation of the fermentation conditions or of the biosynthesis genes, in order to produce desired analogs of the original structure. The availability of genes involved in the formation of secondary metabolites has been exploited for the formation of derivatives of natural products obtained after genetic manipulation of the producing organism (Hopwood, 1997, Chem. Rev. 99:0-39). These manipulations have resulted in novel molecules, many of which would be extremely hard if not impossible to produce by chemical derivatization of the parent compound. The obvious economical and environmental benefits resulting from the formation of the desired structure in one fermentation step constitute an additional stimulus for the application of pathway engineering for the rational design of novel structures. The compounds obtained in this way are amenable evaluation of their biological properties as well as being substrates for further derivatization by chemical or biological means.

In summary, the supply of a natural product produced by a newly discovered microorganism, the optimization of the complex composition, and the process of lead optimization will all benefit from a detailed knowledge of the genetics and physiology of the producing strain. The present invention describes a general method for transferring the capability to

produce any secondary metabolite from the original actinomycete producer to an established and genetically manipulatable production host. The general concept of the invention is illustrated in Fig. 1. Conditions for optimal growth, metabolite production and maintenance need therefore to be developed for one host. In addition, the availability of the cloned genes in a genetically manipulatable and well characterized host allows the utilization of all the genetic tools developed for these strains for the creation of novel derivatives of the natural product after genetic intervention.

3. SUMMARY OF THE INVENTION

The present invention provides a system for producing and manipulating natural products produced by a large group of bacteria for the purpose of drug discovery, development and production. The method of the invention transfers the ability to produce a secondary metabolite from an actinomycete that is the original producer of the natural product, to another production host that has desirable characteristics.

In one embodiment, the invention involves the construction of a library from a donor organism, the producer of a natural product, in an artificial chromosome that can be shuttled between a convenient, neutral cloning host, such as the bacterium *Escherichia coli*, and a production host, such as the actinomycetes *Streptomyces lividans* or *Streptomyces coelicolor*. The clones directing the synthesis of the natural product

are identified in said library, transferred into the production host where said natural product is synthesized.

In another embodiment, the invention involves the reconstruction of a large segment that directs the synthesis of a natural product, starting from smaller DNA fragments cloned from the genome of a donor organism. This reconstruction occurs in an artificial chromosome that can be maintained in a convenient neutral host, such as the bacterium *Escherichia coli*, and subsequently transferred into an actinomycete production host. The reconstructed genomic segment in the artificial chromosome is transferred into the production host where said natural product is synthesized.

The present invention also relates to *Escherichia coli-Streptomyces* Artificial Chromosomes, recombinant DNA vectors useful for shuttling the genetic information necessary to synthesize a given natural product between a donor actinomycete producer and a production host.

3.1 DEFINITIONS

As used herein, the following terms will have the meaning indicated.

An "*Escherichia coli-Streptomyces* Artificial Chromosome" is a recombinant DNA vector that can accept and maintain very large DNA inserts in an *Escherichia coli* host, and that can be introduced and maintained in an actinomycete production host.

A "natural product" is a secondary metabolite made by a microorganism through a series of biosynthetic steps. This natural product may or may not have any useful biological activity.

A "complex" is the mixture of related natural products with similar properties and biological activity that are often produced by the same biosynthetic pathway.

A "donor organism" is the original producer of a natural product, where the synthesis of said compound is governed by a defined number of genetic elements.

A "gene cluster", a "cluster", a "biosynthesis cluster" all designate a contiguous segment of the donor organism's genome that contains all the genes required for the synthesis of a natural product.

A "production host" is a microorganism where the formation of a natural product is directed by a gene cluster derived from a donor organism.

As used in the present invention, the following abbreviations are employed: °C (Celsius degree); h (hour); min (minute); kb (kilobase); µl (microliter); ml (milliliter); mm (millimeter); mg (milligram); µg (microgram); ng (nanogram); M (molar); Mb (megabase); UV (ultraviolet); kV (kilovolt); Ω (Ohm); mFa (millifaraday).

In addition, the following abbreviations are used: Ab, antibiotic; Ap, ampicillin; bp, base pair; ca., circa (i.e. "about"); Cm, chloramphenicol; *E.*, *Escherichia*; ESAC, *E. coli-Streptomyces* Artificial

Chromosome; GC, guanosine + cytosine; HTS, high throughput screening; Km, kanamycin; LB, Luria Broth; LMP, low melting point; *P.*, *Planobispora*; PCR, polymerase chain reaction; PFGE, Pulsed Field Gel Electrophoresis; ^{R}, resistance; rpm, rounds per minute; *S.*, *Streptomyces*; ^{S}, sensitive; *Sac.*, *Saccharopolyspora*; SDS, sodium dodecyl sulfate; Tc, tetracycline; TE, TrisHCl EDTA buffer; Th, thiostrepton; *ts*, temperature sensitive; U, units; vol, volume; wt, weight; YEME, yeast extract malt extract medium.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Scheme of the invention. The general concept of the invention, whereby the gene cluster required for the synthesis of a natural product in a donor organism is established in an ESAC vector in an *Escherichia coli* host, and then transferred into a desired production host, where it integrates into the chromosome and directs production of the secondary metabolite. The hexagon represents the natural product, the twisted thin line the bacterial chromosomes, and the thick line the desired gene cluster. The ESAC episome is represented by a circle.

Figure 2. *E. coli-Streptomyces* artificial chromosome vectors. Vectors pPAC-S1 and pPAC-S2 differ solely for the orientation of the *int-tsr* cassette. Relevant features of the vectors are illustrated. Suitable cloning sites are shown as: B, *Bam*HI; S, *Sca*I; X, *Xba*I. The replicating function of bacteriophage P1

are indicated by the thick bars.

Figure 3. General scheme of the invention, top-down approach. High molecular weight DNA from the donor organism is cloned into an ESAC vector. The resulting library in *E. coli* is screened with the required probes, and the relevant ESAC clones are identified. These are introduced into the desired production host strain, where they integrate site-specifically into the host chromosome. Symbols and abbreviations are as in Fig. 1.

Figure 4. General scheme of the invention, bottom-up approach. A cosmid library is prepared with DNA from the donor organism and screened with the required probes. The overlapping inserts from the positive cosmids are assembled into an ESAC vector via homologous recombination in *E. coli*. The reconstructed ESAC clone is introduced into the desired production host, where it integrates site-specifically into the host chromosome. Symbols and abbreviations are as in Fig. 1.

Figure 5. Scheme of assemblage. The figure illustrates a hypothetical genomic segment from a donor organism that is covered by the inserts from three overlapping clones. The relevant fragments A and D, which denote the ends of the segment, and B and C, which represent regions of overlap, are indicated with their relative orientation (thick side on the fragment rectangle). The bottom part illustrates the reconstructed ESAC clone.

Figure 6. Constructs required for cluster

assemblage. The plasmids indicated are generated by routine in vitro DNA manipulations. Fragments A, B, C and D are as in Fig. 5. Fragment pairs are in this example separated by a marker, indicated as Ab^R for antibiotic resistance. Selective markers present on the two compatible replicons are, as an example: Cm^R and Km^R.

Figure 7. Interplasmid insert exchange. Each of the Cm^R derivatives, as of Fig. 6, is introduced in the same *E. coli* cell as the cognate clone of Fig. 5 (for example a cosmid that carries a Km^R marker). Formation and then resolution of the cointegrate leads to the transfer of the cosmid's insert, indicated here by a looping line, in the Cm^R replicon.

Figure 8. Sequel of assembling steps. A series of interplasmid cointegration and resolution events is conducted. Only the growing ESAC clone is indicated. The starting ESAC clone (Fig. 6) is recombined with plasmid pAB2 (Fig. 7), leading to the insertion into ESAC of the insert flanked by fragments A and B. Next, the Ab^R marker from pBC1 (Fig. 6) is introduced between fragments B and C, and subsequently replaced by the insert from pBC2 (Fig. 7). Finally, the Ab^R marker from pCD1 (Fig. 6) is introduced between fragments C and D, and subsequently replaced by the insert from pCD2 (Fig. 7).

Figure 9. A gene cluster from *Planobispora rosea*. A restriction map of the *Bam*HI sites (indicated as short vertical lines) in a gene cluster from *P. rosea* ATCC 53733 is reported, together with the cosmids pRP16,

pRP31 and pRP58. The fragments A, B, C and D used for assemblage are highlighted. Restriction sites are abbreviated as: M, *Sma*I; P, *Pst*I; S, *Sac*I.

Figure 10. Signature sequences at the left (panel A) and right (panel B) ends of the cluster of Fig. 9. The sequence in panels A starts around coordinate 1.8 kb (Fig. 9); the sequence in panel B ends around coordinate 91.0 kb (Fig. 9). The orientation of the sequences is the same as in Fig. 9.

Figure 11. Site-specific integration of an ESAC clone. PFGE analysis of *S. lividans* ZX7 transformed with ESAC-70. Lanes 1 and 2: *S. coelicolor* M145; lane 3: *S. lividans* ZX7 DNA; lane 4: ZX7 *attB*::ESAC-70 DNA, colony 1; lane 5: ZX7 *attB*::ESAC-70 DNA, colony 2; lane 6: 50-kb ladder, size marker. All DNAs in lanes 1-5 are digested with *Dra*I. Conditions for PFGE are: 200 Volts, 70 s switching for 7 15 h, 120 s switching for 11 h.

Figure 12. Characterization of *S. lividans* transformants. Southern hybridization of *S. lividans* *attB::pPAD6*, grown with (lane 1) or without (lane 2) thiostrepton. *P. rosea* DNA is shown as control (lane 3). Lane 4 contains 1-kb ladder. All DNAs are digested with *Bam*HI and probed with labeled pPAD6.

5. DETAILED DESCRIPTION OF THE INVENTION

In its broadest sense, the present invention entails a general procedure for constructing a *Streptomyces* host producing any natural product after selective transfer of the relevant genes from the original actinomycete producer, the donor strain. This general procedure is

outlined in Fig. 1. The present invention can be applied with only limited information on the structure of the natural product and very little knowledge of the original producer's genetics. The present invention has a substantial impact on the process of drug discovery involving natural products or their structural derivatives. The transfer of the producing capability to a well characterized host can substantially improve several portions of the process of lead optimization and development: the titer of the natural product in the producing strain can be more effectively increased; the purification of the natural product can be carried out in a known background of possible interfering activities; the composition of the complex can be more effectively controlled; altered derivatives of the natural product can be more effectively produced through manipulation of the fermentation conditions or by pathway engineering. In order to better understand the value of the present invention, a brief description is reported below of the current methods for optimizing the productivity of the producing strain, for purifying a natural product, for controlling the composition of a complex, and for producing derivatives of a natural product.

The production of a natural product is controlled by several mechanisms, few of which have been established in detail. Generally, the level of production of a natural product depends on the composition of the growth medium; on the presence of appropriate precursors or on the absence of specific

inhibitors; on the expression level and timing of genes controlling the biosynthetic pathway and competing routes; and on the level and specific activity of key enzymes in the pathway. Because of this complexity, the productivity of the original strain is usually increased by an empirical process, which may include, among other things, one or more of the following steps: strain purification, selection of phenotypic variants arising spontaneously or after mutagenic treatment of the strain, variation in the fermentation medium or in the fermentation parameters; genetic engineering of the producing strain. Fundamental knowledge about the physiology of the producing strain and the variables affecting titer must be achieved for an effective improvement of productivity. This knowledge is very scant in a newly identified producer strain.

During the discovery and development phase, sufficient quantities of a natural product must be available for an evaluation of its properties and/or for the generation of analogs. Because of its uniqueness, a specific purification process must be developed for each natural product. However, it is highly desirable to have the natural product as free as possible of compounds that may interfere with the biological activity of the molecule. Contaminating impurities must be characterized analytically and biologically. In a poorly characterized producer, little information is available on the relevance of contaminating impurities.

A natural product may be produced by a

microorganism as a complex of a few or tens of molecules with minor structural differences, designated congeners. Although most of the congeners are usually biologically active, only one or a few may represent the desired product: for example, one congener may be substantially more active than the others; it may possess better physico-chemical properties; or it may be a better substrate for chemical modification. The composition of a complex can be somehow controlled by intervening on the fermentation parameters. However, the most effective way is usually the altered expression of selected genes by genetic engineering (e.g. Sezonov et al., 1997, Nature Biotechnol. 15:349-353).

Chemical modification of natural products represents the most commonly used means of obtaining novel structures. This approach has been successfully employed, but it still suffers from practical limitations to the number and type of compounds obtainable. The structural complexity of many natural products makes their total synthesis often too lengthy and expensive to be of any practical use. This same structural complexity, with either the presence of several closely related functional groups or their absence, limits modification of a natural product to selected portions of the molecule. Methods of combinatorial synthesis need an initial scaffold as the starting building block, and this can be often generated only through a low yield degradation of the natural product. However, derivatives of natural

products that would be very hard if not impossible to produce by chemical means have been obtained after genetic alteration of the biosynthetic pathway. Examples include the introduction of additional genetic information (Epp et al., 1989, Gene 85:293-301), the targeted inactivation of selected genes or portion thereof (Donadio et al., 1993, Proc. Natl. Acad. Sci. USA 90:7119-7123), the "mixing and matching" of genes or portions thereof from different pathways (McDaniel et al., 1994, Nature 375:549-554).

All the above activities are important for the process of lead optimization and for the development of selected lead structures. They can all benefit, to different extent, from a detailed knowledge of the physiology of the producing strain, and from the possibility of genetically manipulating it. The process by which a given organism is genetically manipulated in order to alter the type, quality or quantity of a natural product is referred to as pathway engineering. The ability to perform pathway engineering in a newly isolated microorganism producing a bioactive molecule with promising characteristics can therefore considerably expedite the optimization of a lead structure and the development process. Pathway engineering can be schematized as a sequel of three steps: a) isolation of the genes of interest; b) performing on selected gene(s) the manipulations required by the specific objective; and c) introduction of the modified gene(s) in suitable form in an appropriate host.

Isolation of the genes of interest from most actinomycetes can be achieved quite easily. The genes for primary metabolism are usually well conserved, and they can be easily accessed in any microorganism by using suitable hybridization probes or by the PCR. The genetic elements governing the biosynthesis of the major classes of secondary metabolites have been also described, and many genes can similarly be identified. Since natural product biosynthesis is governed by clusters, one needs to identify just a few genes in order to have them all. However, synthesis of the vast majority of natural products requires a considerable extent of genetic information. For examples, biosynthesis of the natural products erythromycin (an antibiotic), avermectin (an antihelmintic agent) and rapamycin (an immunosuppressant) requires 55, 90 and 95 kb, respectively, of genetic information (Katz and Donadio, 1993, Annu. Rev. Microbiol. 47:875-912; MacNeil, 1995, Avermectins, in *Genetics and Biochemistry of Antibiotic Production*, Vining and Stuttard eds., Butterworth-Heinemann, Boston CT, p.421-442; Schwecke et al., 1995, Proc. Natl. Acad. Sci. USA 92:7839-7843). Other natural products may require even larger extent of genetic information. Therefore, in order to isolate an entire cluster in a single piece, cloning vectors capable of accepting and maintaining large DNA segments are necessary.

The manipulation of the isolated genes is generally best performed in a convenient cloning host, such as *E. coli*. Manipulations relevant to pathway

engineering can include some or all of the following: site directed mutagenesis, gene inactivation, gene fusions, modification of regulatory sequences, etc. Techniques for the in vitro manipulation of DNA and for the propagation of the mutated alleles in *E. coli* are well developed and can be applied to DNA from virtually any source (Sambrook et al., 1989, In *Molecular Cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

The final step in pathway engineering requires the introduction of modified or heterologous gene(s), in suitable form, in a strain where these genes can be appropriately expressed. This strain is often the strain producing the natural product whose quantity, quality or type one wants to alter. The genes of interest must be carried on appropriate vectors: according to the particular objective of pathway engineering, one may need, among others, vectors that can be stably maintained as single or multicopy episomes; that can insert into the host chromosome at a fixed location; that allow replacement of an endogenous gene with an *in vitro* modified allele; that allow deletion of selected genes from the host chromosome. In addition, for each strain one must have means for introducing heterologous DNA and selecting for its presence. Therefore, in order to genetically manipulate a given producer, one must establish conditions for rendering the bacterial cell capable of receiving incoming DNA; for selecting the incoming DNA; and develop vectors and methodologies for the various types

of manipulations exemplified above. Low- and high copy-number, integrative, non-replicating vectors must be developed with appropriate selection markers. Thus, for each producing strain, specific gene transfer tools and conditions must be developed, starting in most cases from extremely poor knowledge about the microorganism. In addition, techniques developed for one species do not necessarily apply to a new species from the same genus, and often not even to a new strain. It is then no wonder that, among the thousands of strains described as producers of interesting natural products, gene transfer systems have been developed only for a limited number of species, which serve either as model organisms for genetic and physiological studies, or produce a commercially important molecule. The present invention provides tools for the general manipulation of any secondary metabolite pathway, and overcomes the difficulties of developing *ad hoc* conditions for a new producer.

Naive hosts have been shown to produce the appropriate natural product or its intermediate(s) when the relevant DNA was introduced into them (Malpartida and Hopwood, 1984, Nature 309:462-464; Hong et al., 1997, J. Bacteriol. 179:470-476; Kao et al., 1994, Science 265:509-512; McGowan et al., 1996, Mol. Microbiol. 22:415-426; Kealey et al., 1998; Proc. Natl. Acad. Sci. USA 95:505-509). However, the examples reported thus far represent special cases. Indeed, they include the introduction of relatively small DNA segments into a production host; or the transfer of

gene clusters within members of the same bacterial genus; or they have required the careful engineering of specific biosynthesis genes under the control of appropriate genetic elements that direct their expression. Furthermore, the *Streptomyces* vectors currently available have an upper limit of ca. 40 kb (Hopwood et al., 1987, Methods Enzymol. 153:116-167).

Until now, it was not established that DNA fragments exceeding 100 kb, derived from the high GC genome of actinomycetes, could be cloned and stably maintained in an *E. coli* host. Nor was any report of the introduction of large DNA segments into a *Streptomyces* host. The unexpected finding described herein is that these cloning tasks can be achieved according to the principles and methodologies of the present invention. Furthermore, the genetic elements required for the synthesis of a natural product in the original producer are genetically stable in a heterologous host, where they can direct the synthesis of the desired molecule. It was also unexpected and unprecedented that this heterologous stability and expression can occur when the donor organism and the production host belong to different bacterial genera.

The present invention rests on the fact that the genes required for the formation of a natural product are found as gene clusters of a defined size; that these gene clusters can be conveniently isolated, manipulated and transferred among different actinomycete strains; that they are expressed in a heterologous host; and on the fact that all the primary

metabolite precursors required for the formation of a particular natural product are either produced by selected enzymes encoded by cluster-specific genes, or are present and available in the heterologous host at the time of formation of the natural product. The present invention addresses also the crucial aspect of natural product formation in actinomycetes: i.e. synthesis of many natural products may require over 100 kb of genetic information. To be generally applicable, transferring all the genes necessary for the production of any natural product requires cloning vectors capable of accommodating fragments as large as 150 kb, and possibly more. An object of the present invention is therefore represented by vectors capable of accommodating such large fragments which are also capable of being stably maintained in a suitable microbial host, such as a *Streptomyces* host.

Examples of these vectors are designated with the generic name ESAC: *E. coli-Streptomyces* Artificial Chromosomes. They are derived from bacterial artificial chromosomes (Shizuya et al., 1992, Proc. Natl. Acad. Sci. USA 89:8794-8797; Ioannou et al., 1994, Nature Genet. 6:84-89)and can carry inserts up to 300 kb, or more.

As a general example of the broad applicability of the principles and methodologies described in the present invention, the Examples reported below describe how a convenient *Streptomyces* host can be engineered to carry a large gene cluster in order to produce a desired natural product through the use of an

appropriate ESAC vector. The exemplary organism chosen as the donor organism is the actinomycete *P. rosea*, belonging to one of the lesser characterized genera of actinomycetes (Goodfellow, 1992, In *The Prokaryotes*, 2nd edn., Balows, Trueper, Dworkin, Harder and Schleifer eds, Springer-Verlag, New York, NY, USA). This organism produces the natural product GE2270 (Selva et al., 1991, J. Antibiotics 44:693-701), an antibacterial agent. This particular case therefore describes the general applicability of the present invention, since very little information is available on the donor organism, on its genetics and physiology, and on the gene clusters present in its genome. Further examples described herein illustrate the application of the principles and methodologies of the present invention to other gene clusters described in the literature.

The present invention, relating to a general methods for transferring the capability to produce any natural product from the original actinomycete to an established and genetically manipulatable *Streptomyces* host, can be schematized in a series of passages summarized as: 1) design of suitable vectors; 2) construction of a large-insert library in said vectors; 3) selection of the desired clones with appropriate probes; 4) insertion of the selected clones into a convenient *Streptomyces* host; and 5) growth of the recombinant strain under appropriate conditions to produce the natural product.

Actinomycetes produce a large number of natural

products with important applications. However, other important classes of microbial producers are known, and newer ones are likely to be discovered in the upcoming years, as more microbial sources are screened for potential new drugs. Important classes of microbial producers include, among others, filamentous fungi, bacilli, mixobacteria, pseudomonas and cyanobacteria. The series of passages described above can therefore be applied to other important classes of microbial producers, provided that two requisites are met: the synthesis of the desired natural product is governed by a gene cluster; suitable production host(s) exist; and appropriate selective marker(s) and maintenance function(s) are introduced into the artificial chromosome. Therefore, the principles and methodologies described in the present invention can be extended to microbial producers other than actinomycetes.

Furthermore, the series of passages summarized above and described in detail in the Examples, involve the use of a neutral cloning host. This host, as described in the present invention, is the bacterium *Escherichia coli*. In a preferred example of such a host, a high cloning efficiency can be obtained, and many of the analyses of the ESAC clones can be quickly performed. However, it is evident to one of ordinary skill in this art that any other host that allows high cloning efficiency can be used as neutral cloning host. Additionally, the use of such a host is not a *conditio sine qua non* for the applicability of the present invention. In fact, when it is possible to establish

directly a library in a production host, there is no need for an intermediate neutral cloning host.

6. GENERAL METHODS

Plasmids, Bacterial Strains and Growth Conditions

Plasmids pUCBM20, pUCBM21, pBR322 and pUC18 are obtained from Boheringer Mannheim; plasmid pIJ39 and ΦC31 DNA are from prof. David Hopwood, The John Innes Centre, Norwich, UK; plasmid pCYPAC2 is from prof. Pieter de Jong, Roswell Park Cancer Institute, Buffalo, NY, USA; plasmid pMAK705 from prof. Sidney Kushner, University of Georgia, Athens, USA; cosmid Lorist6 from prof. Stewart Cole, Pasteur Institute, Paris, France. *E. coli* strains DH5α, DH10B, C600, DH1 and XL1blue are obtained from commercial sources. *S. coelicolor* M145 and *S. lividans* ZX7 are from prof. David Hopwood, The John Innes Institute, Norwich, UK. *Planobispora rosea* ATCC 53733, *Streptomyces hygroscopicus* ATCC 29253, *Saccharopolyspora erythraea* NRRL2338 are from the ATCC culture collection. *Amycolatopsis mediterranei* S699 has been described (August et al., 1998, Chem. Biol. 5:69-79). All other materials are from commercial sources. Media for cultivation of *E. coli* (Sambrook et al., 1989, In *Molecular Cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press) and *Streptomyces* (Hopwood et al., 1985, *Genetic Manipulation of Streptomyces: A Laboratory Manual*, The John Innes Foundation, Norwich, UK) have been described. The JM medium for *S. coelicolor* has been described (Puglia et al., 1995,

Mol. Microbiol. 17:737-746).

DNA Manipulations DNA manipulations are performed following described procedures, using the appropriate *E. coli* strains as cloning hosts (Sambrook et al., 1989, In *Molecular Cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press). Genomic DNA from actinomycetes is prepared as described (Hopwood et al., 1985, *Genetic Manipulation of Streptomyces: A Laboratory Manual*, The John Innes Foundation, Norwich, UK). A cosmid library of *P. rosea* DNA is constructed in the cosmid vector Lorist6 following published procedures (Sambrook et al., 1989, In *Molecular Cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press). Amplification by the PCR are performed following published guidelines (Innis, Gelfand, Sninsky and White, eds., 1990, PCR Protocols: A guide to Methods and Applications, Academic Press, San Diego, CA, USA).

Hybridizations Probes Pep6 and Pep8 are derived from conserved motifs in peptide synthetase gene sequences (Turgay and Marahiel, 1994, Pept. Res. 7:238-241). Oligonucleotide probe Pep6 consists of an equimolar mixture of 5'-GCSTACATCATCTACACSTCSGGSACSACS-GGSAAGCCSAAGGG-3' and 5'-GGSTACATCATCTACACSAGCGGSACSAC-SGGSAAGCCSAAGGG-3'. Oligonucleotide probe Pep8 consists of an equimolar mixture of 5'-AKGCTGTCSCCSCCSAGSNNGAAG-AAGTYGTCGTCGATSCC-3' and 5'-AKGGAGTCSCCSCCSAGSNNGAAGAA-GTYGTCGTCGATSCC-3'. [*S* indicates G or C; *K* indicates G or T; *Y*, C or T; and *N*, any base]. Hybridizations are

performed with a hybridization stringency set at 2xSSC, 55 °C, and a final wash set at the same stringency.

Preparation of high molecular weight DNA Procedures for the preparation of high molecular weight DNA from actinomycetes for PFGE have been described (Dyson, 1993, Trends Genet. 9:72; Kieser et al., 1992, J. Bacteriol. 174:5496-5507). They are modified for constructing libraries as described in the Examples.

7. EXAMPLES

The present invention consists in a series of passages, involving the design of suitable vectors; the introduction of large DNA inserts in said vectors employing genomic DNA from the donor organism; the selection of clones carrying the cluster specifying the synthesis of the desired natural product; the introduction of selected clone(s) into the appropriate production host; and the growth of the recombinant strain under appropriate conditions for metabolite production. These passages are described in detail in the Examples reported herein. These Examples outline the steps necessary to accomplish each passage, for the overall purpose of the present invention: the production of a natural product in a different host. They serve to illustrate the principles and methodologies of the present invention, and are not meant to restrict its scope to the Examples specified herein.

7.1 Cloning vectors

Bacterial artificial chromosomes are circular plasmids that can be easily propagated in and prepared from *E. coli* cells by standard miniprep methods (Shizuya et al., 1992, Proc. Natl. Acad. Sci. USA 89:8794-8797; Ioannou et al., 1994, Nature Genet. 6:84-89). In order to adapt bacterial artificial chromosomes to a *Streptomyces* host, they need to be endowed with a selectable marker and maintenance functions. Site-specific integration, mediated by the action of an integrase encoded by the *int* gene, allows the stable incorporation of episomal elements into the host genome, at a defined locus designated *attB*. The episomal element needs to carry the cognate *attP* site and it may lack replicative functions. In addition, *int*-mediated excision of the integrated element from the chromosome *via* reversal of the integration event can be prevented through selection of the resistance marker carried by the integrated episome; or, if necessary, after site-specific integration has occurred, the *int* gene on the integrated episome can be inactivated. Site-specific integration therefore allows the introduction of foreign DNA in single copy at a defined genetic locus. Several systems capable of directing site-specific integration of incoming circular DNA into the chromosome of a *Streptomyces* host have been described. A convenient system that can be used in the present invention is for istance the *int-attP* system derived from the temperate bacteriophage ΦC31 (Kuhstoss and Rao, 1991, J. Mol. Biol. 222:897-908), which directs, during lysogen formation,

integration of the 41-kb phage genome at the *attB* site, located in a stable segment of the *S. coelicolor* chromosome (Redenbach et al., 1996, Mol. Microbiol. 21:77-96). Several selectable markers have been described that can be used for *Streptomyces* (Hopwood et al., 1985, Genetic Manipulation of *Streptomyces*: A Laboratory Manual, The John Innes Foundation, Norwich, UK). The *tsr* gene, conferring resistance to the antibiotic thiostrepton (Thompson et al., 1982, Gene 20:51-62), is used in the present invention. The ESAC vectors pPAC-S1 and pPAC-S2, described in the present invention, are depicted in Fig. 2. Their relevant features are: ability to accommodate DNA inserts up to 300 kb; low copy number in *E. coli* for increased stability; ease of propagation in *E. coli* because of the pUC19 stuffer segment; *Bam*HI, *Xba*I or *Sca*I cloning sites, with positive selection of inserts for resistance to sucrose; T7 and SP6 promoters flanking the cloning site; Km^R or Th^R for selection in *E. coli* or actinomycetes, respectively; site-specific integration at the ΦC31 *attB* site into the *Streptomyces* genome. Vectors pPAC-S1 and pPAC-S2 are 22 kb in size and differ solely for the orientation of the *int-tsr* cassette. After release of the stuffer pUC19 segment, the vector size is reduced to 19.7 kb. When cloning in the *Bam*HI site, the vector can be released by digestion with *Dra*I, resulting in vector fragments of 7.4, 4.2 and 0.6 kb. The additional 7.5 kb of vector DNA will be associated with the insert. *Dra*I rarely cuts in the high-GC genome of actinomycetes, so that

the insert size can be easily calculated.

Example 1

Isolation of the *int* region from ΦC31

Two pairs of PCR primers, 5'-TTTTTGGTACCTGACGTCCCGAAGG-CGTG-3' and 5'-CAGCTTGTCCATGGCGGA-3'; and 5'-TCTGTCCGC-CATGGACAAGC-3' and 5'-TTTTTGGATCCGGCTAACTAACTAAACCGAGA-3', are used to amplify the *int*-containing fragments of 1.3 and 0.9 kb, respectively. The template is ΦC31 DNA. The amplified fragments are digested with *Kpn*I + *Nco*I and *Nco*I + *Bam*HI, respectively, and recovered from an agarose gel.

Example 2

Construction of plasmid pINT

The 1.3 and 0.9 kb fragment, prepared as described in Example 1, are ligated to pUCMB21, digested with *Kpn*I + *Bam*HI. The resulting mixture contains the desired plasmid pINT.

Example 3

Construction of *E. coli* K12 DH5α/pINT

Approximately 10 ng of plasmid pINT, prepared as described in Example 2, are used to transform *E. coli* DH5α and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pINT, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 4.0 and 0.9 kb after digestion of the plasmid with *Nco*I + *Bam*HI.

Example 4

Construction of plasmids pUIT1

The 1.8 kb *Bam*HI fragment containing the *tsr* gene is isolated from pIJ39 and ligated to pINT, prepared as described in Example 3 and previously digested with *Bam*HI. The resulting mixture contains the desired plasmids pUIT1.

Example 5

Construction of *E. coli* K12 DH5α/pUIT1

Approximately 10 ng of plasmid pUIT1, prepared as described in Example 4, are used to transform *E. coli* DH5α and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUIT1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 4.9 and 1.8 kb after *Bam*HI digestion of the plasmid.

Example 6

Construction of plasmid pUIT3

The 3.7 kb *Apa*I fragment, containing the *int-tsr* cassette, is isolated from plasmid pUIT1, prepared as described in Example 5, and ligated to pUCBM21 digested with *Apa*I. The resulting mixture contains the desired plasmid pUIT3.

Example 7

Construction of *E. coli* K12 DH5α/pUIT3

Approximately 10 ng of plasmid pUIT3, prepared as described in Example 6, are used to transform *E. coli* DH5α and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUIT3, as verified by the observation, upon agarose gel-electrophoresis, of fragments of of 4.2 and 2.2 kb after *Bam*HI digestion of the plasmid.

Example 8

Construction of plasmid pUIT4

The *Bam*HI site present in the *int-tsr* cassette of plasmid pUIT3 is eliminated as follows. Plasmid pUIT3, prepared as described in Example 7, is partially digested with *Bam*HI, followed by filling-in of the resulting ends, and treated with DNA ligase. The resulting mixture contains the desired plasmid pUIT4.

Example 9

Construction of *E. coli* K12 DH5α/pUIT4

Approximately 10 ng of plasmid pUIT4, prepared as described in Example 8, are used to transform *E. coli* DH5α and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUIT4, as verified by the observation, upon agarose gel-electrophoresis, of a 6.4 kb fragment after *Bam*HI digestion of the plasmid.

Example 10

Construction of plasmid pPAC-S1 and pPAC-S2

The 3.7 kb *Apa*I fragment from pUIT4, prepared as described in Example 9, is mixed with pCYPAC2, previously digested with *Nhe*I. After filling-in of the ends, DNA ligase is added. The resulting mixture contains the desired plasmids pPAC-S1 and pPAC-S2.

Example 11

Construction of *E. coli* K12 DH10B/pPAC-S1 and DH10B/pPAC-S2

Approximately 10 ng of plasmids pPAC-S1 and pPAC-S2, prepared as described in Example 10, are used to transform *E. coli* DH10B and a few of the resulting Km^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pPAC-S1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 8.1, 4.8, 4.6, 2.2, 2.2, 0.5 and 0.1 kb after *Eco*RI digestion of the plasmid. Another colony is found to carry pPAC-S2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 8.1, 7.8, 2.2, 2.2, 1.5, 0.5 and 0.1 kb after *Eco*RI + *Bam*HI digestion of the plasmid.

Although the present invention is described in the Examples listed above in terms of preferred embodiments, they are not to be regarded as limiting the scope of the invention. The above Examples serve to illustrate the principles and methodologies for constructing bacterial artificial chromosomes that can

be introduced in a *Streptomyces* host. It will occur to those skilled in the art that selectable markers different from the *tsr* gene can be employed for selection in *Streptomyces*. Other useful markers are described in detail in laboratory manuals (Hopwood et al., 1985, Genetic Manipulation of *Streptomyces*: A Laboratory Manual, The John Innes Foundation, Norwich, UK) and include but are not limited to: genes conferring resistance to apramycin, kanamycin, erythromycin, hygromycin, viomycin. It will also occur to those skilled in the art that functions other than those specified by ΦC31 can be used for directing site-specific integration in the *Streptomyces* chromosome. These functions are described in recent literature (Hopwood and Kieser, 1991, Methods Enzymol. 204:430- 458) and include but are not limited to those derived from pSAM2, SLP1, IS117. Bacterial artificial chromosomes derived from the *E. coli* F plasmid have been described (Shizuya et al., 1992, Proc. Natl. Acad. Sci. USA 89:8794-8797). It will occur to those skilled in the art that, using the principles and methodologies described above, the *int-tsr* cassette from pUIT4, prepared as described in Example 9, could be inserted into a unique site of pBAC108L (Shizuya et al., 1992, Proc. Natl. Acad. Sci. USA 89:8794-8797) or of suitable derivatives of this vector, leading to the formation of a BAC-based series of ESAC vectors. It will occur to those skilled in the art that other ESAC vectors differing, for example, in their size, in the *E. coli* replicon they carry, in the selectable marker for *E.*

coli, in the cloning sites, can also be used in the present invention. Other differences and variations in the technical aspects of the present invention could be employed. These include but are not limited to: different methods and sources for obtaining selectable markers and integrative functions; different cloning sites and methodologies; different *E. coli* hosts for amplifying the recombinant constructs. All these variations fall within the scope of the present invention.

7.2 Construction of large inserts in ESAC

Two distinct methodologies for introducing large DNA fragments into the vectors described in Section 7.1 fall within the scope of the present invention. The first methodology can be referred to as the top-down approach and is depicted in Fig. 3. It consists of directly cloning the desired gene cluster into an ESAC vector through the construction of a genomic library of DNA fragments of average size of 100 kb, or more. The library is then screened with suitable probes (Section 7.3) in order to identify the desired cluster. The second methodology can be considered a bottom-up approach and is illustrated in Fig. 4. It consists of assembling the desired gene cluster from pre-existing smaller segments of cloned, overlapping DNA, through the iterative use of homologous recombination in *E. coli*. The desired overlapping clones encompass the desired gene cluster and are identified as described in Section 7.3. Both methodologies fall within the scope

of this invention. Depending on factors such as previous knowledge about the biosynthesis cluster, the extent of characterization of the producing strain, the existence of other natural products of interest produced by the original microorganism, one methodology may be preferred over the other. However, the two methodologies are not mutually exclusive.

7.2.1 Preparation of a large insert library

In order to prepare a large-insert library, particular care must be taken in the preparation of genomic DNA from the actinomycete strain of choice. Although several procedures have been described for the isolation of genomic DNA, few are suitable for obtaining sufficient yields of high molecular weight DNA. The strain of choice is grown in a medium that allows dispersed growth to facilitate lysis of the cells. Examples of suitable growth media for different genera of actinomycetes can be found in the literature (Balows, Trueper, Dworkin, Harder and Schleifer eds., 1992, *The Prokaryotes*, 2nd edn., Springer-Verlag, New York, NY, USA). The growth time should allow formation of a sufficient quantity of biomass; however, long incubation times should be avoided, since mycelia are generally more resistant to lysis as they age. The mycelium is pelleted, washed and embedded in agarose for the subsequent lytic steps. Lysis of the cells is achieved by a combination of enzymatic (e.g., incubation with lysozyme and/or achromopeptidase) and mild physical treatments (e.g., SDS). The

concentrations of reagents and the incubation times need to be optimized for each strain. A good starting point is represented by the conditions described in Example 12. The quality of the DNA preparation is checked by PFGE under appropriate conditions. Once a suitable preparation is obtained, the DNA can be digested as described in Example 13. The exact incubation time and the units of restriction endonuclease are adjusted to the particular DNA preparation for optimizing the size and yield of the bulk of digested DNA, which should exceed 150 kb. The partially digested DNA is size-fractionated on a PFGE gel, without exposure to ethidium bromide or UV light, in order to avoid damage to the DNA. The gel slice containing the desired DNA fraction is localized by staining the marker-containing portion of the gel and cut. All subsequent manipulations are performed with great care (Birren and Lai, 1993, Pulsed Field Gel Electrophoresis: A Practical Guide, Academic Press, New York, NY). The size-selected DNA is ligated to an appropriately prepared ESAC vector (see Example 14) employing a high molar excess of vector to insert (ca. 10:1) in order to minimize the formation of chimeric clones (i.e. those constituted by the religation of two uncontiguous inserts). Subsequent steps are performed using published procedures for the cloning in bacterial artificial chromosomes, as described in Examples 16 and 17.

The genome size of actinomycetes is around 8 Mb. Consequently, a 10-genome equivalents library

consisting of 800 clones with an average insert size of 100 kb has >99.9% probability of containing the desired clone (Sambrook et al., 1989, In *Molecular Cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press). Therefore, the average clone in the library will have a 10-kb segment (8,000 kb divided by 800 clones = 10 kb/clone) of unique DNA , i.e. DNA not found in any other clone. Consequently, a 90 kb cluster will have a high chance of being exactly contained within one or two 100-kb clones in a 800-clone library. The number of clones to be screened and the average insert size to be looked for in the ESAC library depends on the expected size of the biosynthesis gene cluster. The larger the difference between the average insert size and the expected size of the gene cluster, the smaller the number of clones to screen in order to identify an entire gene cluster in a single clone. ESAC DNA is prepared from a representative number of clones obtained after electroporation of a ligation mixture and analyzed for determining the frequency of insert-carrying clones and their average size. If necessary, all insert containing clones can be analyzed by miniprep procedure (Birren and Lai, 1993, *Pulsed Field Gel Electrophoresis: A Practical Guide*, Academic Press, New York, NY, USA) and clones carrying inserts below a certain threshold can be discarded. Alternatively, the number of clones carrying insert of the appropriate size can be estimated after analysis of a representative number of ESAC clones. The quality of

the library can be evaluated by probing with cloned genes from the strain (if available), or from highly conserved "housekeeping" genes from a strain with a similar GC content, such as *S. coelicolor*.

Example 12

Preparation of high molecular weight chromosomal DNA

S. coelicolor strain M145 is grown in YEME medium containing 0.5% (wt/vol) glycine for 40 h at 30°C on an orbital shaker (ca. 200 rpm). The mycelium is pelleted by centrifugation, washed with 10.3% sucrose and the chromosomal DNA is extracted from the mycelium embedded in 0.75% LMP agarose by treatment with 1 mg/ml lysozyme and with 1 mg/ml proteinase K in 0.1% SDS for 40 h at 50°C.

Example 13

Preparation of partially digested chromosomal DNA

S. coelicolor M145 chromosomal DNA, prepared as described in Example 12 and embedded in LMP agarose plugs, is partially digested by limiting the magnesium concentration for 20 min with 4 U of *Sau*3AI. The resulting DNA fragments are resolved by PFGE and the size-selected genomic DNA fraction (larger than 100 kb) is recovered and released from the agarose gel by digestion with gelase.

Example 14

Preparation of pPAC-S1 for library construction

The vector pPAC-S1, prepared as described in Example

11, is cut with *Sca*I and then treated with calf intestinal phosphatase. The recovered DNA is then digested with *Bam*HI and treated with an excess of calf intestinal phosphatase. The short *Sca*I-*Bam*HI linker fragments are removed by spin dialysis.

Example 15

Construction of the ESAC library

Size selected genomic DNA, prepared as described in Example 13, is ligated to pPAC-S1, prepared as described in Example 14, employing 300 Molecular Biology Units of T4 DNA ligase in a 50 µl final volume and using a ca. 10:1 molar ratio of vector to insert. The resulting ligation mixture contains the desired ESAC library, consisting of fragments *S. coelicolor* DNA inserted into the pPAC-S1 vector.

Example 16

Introduction of the library into *E. coli* K12 DH10B

The ligation mixture, prepared as described in Example 15, is drop-dialyzed against 0.5 X TE for 2 h using 0.025 mm type VS membranes (Millipore) and a few µl are used to electroporate 40 µl of electrocompetent *E. coli* DH10B cells. The electroporation conditions are: 2.5 kV, 100 Ω and 25 mFa employing the Biorad Gene Pulser II. The cells are plated on LB-agar plates containing 25 µg/ml Km and 5% sucrose to select for recombinant cells harboring insert-carrying pPAC-S1. Individual colonies are picked into 0.1 ml of LB broth containing 25 µg/ml Km in 96-well microtiter plates, where they

are stored at -80 °C after overnight incubation and addition of glycerol to 20% (v/v).

Example 17

Preparation of recombinant ESAC clones

Individual colonies, prepared as described in Example 16, are inoculated into 5 ml of LB broth containing 25 µg/ml Km and grown overnight. ESAC DNA is isolated using the alkaline extraction procedure (Sambrook et al., 1989, In *Molecular Cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press)) without the phenol extraction step. The DNA is analyzed, after digestion with *Dra*I, by PFGE. Three bands of 7.4, 4.2 and 0.6 kb are common to all clones and represent vector DNA. The insert size in the recombinant ESAC clones is calculated by summing up the sizes of the additional *Dra*I fragments and subtracting from this number 7.5 kb, the amount of the pPAC-S1 vector not included with the three *Dra*I fragments.

The examples described above illustrate the principles and methodologies of constructing a large-insert library of *S. coelicolor* DNA in an ESAC vector. Although the present invention is described in the Examples listed above in terms of preferred embodiments, they are not to be regarded as limiting the scope of the invention. The above descriptions serve to illustrate the principles and methodologies for constructing a large-insert DNA library in an ESAC

vector. It will occur to those skilled in the art that other *Streptomyces* strains can be used as a source of DNA for constructing the library. For example, an ESAC library of the rapamycin producer *Streptomyces hygroscopicus* ATCC 29253 can be constructed, employing the procedures reported for PFGE analysis (Ruan et al., 1997, Gene 203:1-9) and applying the principles and methodologies described in Examples 12 through 17.

It will also occur to those skilled in the art that strains from actinomycete genera other than *Streptomyces* can be used as a source of DNA for constructing an ESAC library. These strains can belong to any genus of the order *Actinomycetales*, which include but are not limited to the genera reported in Table 1. As another example, an ESAC library of the erythromycin producer *Saccharopolyspora erythraea* can be constructed, employing the procedures reported for PFGE analysis (Reeves et al., 1998, Microbiology 144:2151-2159) and applying the principles and methodologies described in Examples 12 through 17. Those skilled in the art understand that bacterial taxonomy is a rapidly evolving field and new genera may be described while old genera may be reclassified. Therefore, the list of bacteria genera related to actinomycetes is very likely to change. Nonetheless, the principles and methodologies of the present invention can be applied to any donor organism related to the actinomycetes.

It will also occur to those skilled in the art that different actinomycete strains will require growth

media different from those reported in Example 12. Furthermore, alternative media and conditions for growth can be employed for obtaining mycelia for DNA preparation; that alternative methods of lysis of mycelia can be utilized; that restriction endonucleases other than *Sau*3AI can be equally effective for constructing a library; that other methods for fragmenting DNA can be employed. In addition, it will occur to those skilled in the art that ESAC vectors other than pPAC-S1, which include but are not limited to the possible vectors described in Section 7.1, can be used for constructing a library. Alternative methods for ligating DNA, for introducing the library in *E. coli* cells, and hosts other than DH10B are well described in the literature and can be employed in the present invention. All the above variations in strains,

Table 1

List of exemplary genera of *Actinomycetales* (http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy)

Acidothermus	*Dietzia*	*Phenylobacterium*
Actinobispora	*Elytrosporangium*	*Pilimelia*
Actinocorallia	*Excellospora*	*Pimelobacter*
Actinokineospora	*Exiguobacterium*	*Planobispora*
Actinomadura	*Frankia*	*Planomonospora*
Actinomyces	*Friedmanniella*	*Planopolyspora*
Actinoplanes	*Gardnerella*	*Planotetraspora*
Actinopolyspora	*Geodermatophilus*	*Prauseria*
Actinopycnidium	*Glycomyces*	*Promicromonospora*
Actinosporangim	*Gordona*	*Propionibacterium*
Actinosynnema	*Herbidospora*	*Propioniferax*
Aeromicrobium	*Intrasporangium*	*Pseudonocardia*
Agrococcus	*Janibacter*	*Rarobacter*
Agromyces	*Jonesia*	*Rathayibacter*
Ampullariella	*Kibdelosporangium*	*Renibacterium*
Amycolata	*Kineococcus*	*Rhodococcus*
Amycolatopsis	*Kineosporia*	*Rothia*
Arcanobacterium	*Kitasatoa*	*Rubrobacter*
Arthrobacter	*Kitasatosporia*	*Saccharomonospora*

Atopobium
Aureobacterium
Bifidobacterium
Blastococcus
Bogoriella
Brachybacterium
Brevibacterium
Catellatospora
Catenuloplanes
Cellulomonas
Chainia
Clavibacter
Coriobacterium
Corynebacterium
Couchioplanes
Cryobacterium
Curtobacterium
Dactylosporangium
Demetria
Dermabacter
Dermacoccus
Dermatophilus
Kocuria
Kutzneria
Kytococcus
Lentzea
Luteococcus
Microbacterium
Microbispora
Micrococcus
Microellobosporia
Microlunatus
Micromonospora
Microsphaera
Microtetraspora
Microthrix
Mobiluncus
Mycobacterium
Nesterenkonia
Nocardia
Nocardioides
Nocardiopsis
Oerskovia
Pelczaria
Saccharopolyspora
Saccharothrix
Sanguibacter
Skermania
Spirilliplanes
Spirillospora
Sporichthya
Stomatococcus
Streptoalloteichus
Streptomyces
Streptosporangium
Streptoverticillium
Terrabacter
Terracoccus
Thermoactinomyces
Thermocrispum
Thermomonospora
Tropheryma
Tsukamurella
Turicella

reagents and methodologies that can be employed for preparing a large-insert library of actinomycete DNA into an ESAC vector fall within the scope of the present invention.

7.2.2 Assemblage by homologous recombination

The bottom-up strategy of assembling large fragment from pre-existing smaller segments of cloned DNA is described in this section. This methodology makes use of the same ESAC vectors described in the present invention under Section 7.1. The desired cluster is assembled from existing overlapping clones by the iterative use of homologous recombination in *E. coli*. In the example of Fig. 5, three overlapping clones, designated 1, 2 and 3, and derived from the genome of a donor organism, encompass the desired biosynthesis

cluster. These clones include leftward fragment "A" unique to clone 1; fragment "B" common to clones 1 and 2; fragment "C" common to clones 2 and 3; and rightward fragment "D" unique to clone 3. These fragments can range from a few hundred bp to a few kb, and are thus amenable to routine in vitro DNA manipulations. The number of overlapping clones encompassing the cluster may vary. However, if *n* is the number of overlapping clones that cover the desired genomic segment, the number of fragments to consider will be equal to *n* + *1*. In the example illustrated in Fig. 5, four fragments are required. The cluster of Fig. 5 is reconstructed into an ESAC vector through the use of successive rounds of homologous recombination in *E. coli*. Fragments A and B are cloned in a *ts* vector, as shown in Fig. 6, which carries a selectable marker, Cm^R as exemplified in Fig. 6. The same is done with fragment pairs B-C and C-D (Fig. 6). The relative orientation of each fragment pair in the *ts* vector must be the same as in the gene cluster. The fragments in each pair may be separated by a selectable marker, designated Ab^R in Fig. 6, to monitor interplasmid insert exchange. Therefore, three constructs in the *ts* vector, designated pAB1, pAB2 and pAB3, are required. The A-B-C-D four-fragment cassette is cloned in an ESAC vector (Fig. 6). The relative orientation of the four fragments in the ESAC vector must be the same as in the gene cluster. Again, a selectable marker may separate any of two fragments to monitor interplasmid insert exchange. The original clone (for example, a cosmid,

which carries a selectable marker, Km^R as exemplified in Fig. 7) containing part of the cluster and the cognate *ts* construct (Fig. 6) are introduced into the same *E. coli* cell. The interplasmid cointegrate between the original clone and the *ts* construct is selected at the non-permissive temperature for the *ts* replicon. This occurs through a single, reciprocal homologous recombination mediated by either one of the two fragments in the A-B, B-C or C-D pairs. The cointegrate is then resolved at the permissive temperature, leading to insert exchange between the two replicons (Fig. 7). The presence in the *ts* replicon of the genomic segment comprised between fragments A and B can be monitored by the appearance of Cm^R Ab^S colonies. This is done for clone 1 and pAB1, resulting in pAB2; for clone 2 and pBC1, resulting in pBC2; and for clone 3 and pCD1, resulting in pCD2. Each insert from the original overlapping clones (Fig. 5) is thus transferred into the *ts* replicon, as outlined in Fig. 7. Subsequently, the inserts from clone 1, now present in the *ts* plasmid pAB2, is introduced into the ESAC construct carrying the entire A-B-C-D cassette. This is done by selecting for the interplasmid cointegrate between the pAB2 and the ESAC construct at the non-permissive temperature, and then resolving the cointegrate at the permissive temperature, selecting for Km^R Ab^S colonies. This leads to insert exchange between the two replicons (as shown in Fig. 8). Next, a selectable marker is introduced in the growing ESAC clone between the next fragment pair, again through the use of two rounds of single,

reciprocal homologous recombination mediated by plasmid pBC1, leading to the appearance of Km^R Ab^R colonies. Subsequently, the interplasmid exchange with pBC2 leads to the introduction of the genomic segment comprised between fragments B and C. Finally, the use of pCD1 first and subsequently of pCD2 leads to the reconstruction of the genomic segment into the ESAC vector. Therefore, through the use of alternating steps, the Ab^R marker first and the genomic segment later are introduced between any fragment pair, as schematized in Fig. 8. This iterative procedure results in the reconstruction of the original chromosomal region in the ESAC vector.

A series of examples described herein illustrate how a 90-kb gene segment from the actinomycete *P. rosea* is assembled from three pre-existing cosmids via homologous recombination. The cosmids, designated pRP16, pRP31 and pRP58, are identified in a cosmid library constructed in the vector Lorist6 by the use of selective hybridization probes. The relevant information about the cluster is reported in Fig. 9. The reconstruction of the cluster results in the formation of the intermediate ESAC derivatives pPAD1, pPAD2, pPAD4 and pPAD6, carrying inserts of 10, 39, 68 and 89 kb, respectively. The examples reported herein serve to illustrate the principles and methodologies of the present invention and are not meant to restrict its scope.

Example 18

Isolation of cosmid clones pRP16, pRP31 and pRP58

A cosmid library of *P. rosea* DNA prepared in the vector Lorist6 is screened with oligonucleotide probes Pep6 and Pep8, according to the conditions described under Section 6. Among the positive colonies identified, several cosmids were found to span the ca. 90 kb segment of the *P. rosea* chromosome reported in Fig. 9. Signature sequences close to the left and right ends of this segment are reported in Fig. 10A and 10B, respectively. Three cosmids are chosen for further studies. Cosmids pRP16, pRP31 and pRP58 exhibits, after digestion with *Bam*HI and resolution by agarose gel-electrophoresis, fragments of 7.5, 7.2, 5.6, 5.2, 2.7, 2.0, 1.9, 1.9, 1.8, 1.6, 1.4, 0.9 and 0.7 kb; of 10.5, 6.2, 3.1, 2.8, 2.6, 2.5, 2.1, 1.9, 1.9, 1.5, 1.4, 1.2, 1.0, 1.0, 0.9, 0.9, 0.7, 0.6, 0.5, 0.5, 0.1 and 0.1 kb; and of 10.0, 7.6, 6.7, 6.2, 3.4, 3.0, 2.8, 2.1, 1.0, 1.0, 0.9, 0.9, 0.7, 0.6, 0.5, 0.5, 0.3 and 0.1 kb; respectively.

Example 19

Construction of plasmid pUA1

The 0.9 kb *Sma*I-*Sst*I fragment, comprised between map coordinates 2.0-2.9 kb of Fig. 9, is obtained from cosmid pRP16, prepared as described in Example 18, and ligated to pUC18 previously digested with *Sst*I and *Sma*I. The resulting mixture contains the desired plasmid pUA1.

Example 20

Construction of *E. coli* K12 XL1blue/pUA1

Approximately 10 ng of plasmid pUA1, prepared as described in Example 19, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUA1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 0.9 kb after digestion of the plasmid with *Bam*HI + *Sst*I.

Example 21

Construction of plasmid pUA2

The 0.9 kb *Bam*HI-*Sst*I fragment from pUA1, prepared as described in Example 20, is ligated to pUCBM20 previously digested with *Bam*HI and *Sst*I. The resulting mixture contains the desired plasmid pUA2.

Example 22

Construction of *E. coli* K12 XL1blue/pUA2

Approximately 10 ng of plasmid pUA2, prepared as described in Example 21, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUA2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 0.9 kb after digestion of the plasmid with *Eco*RI + *Sst*I.

Example 23

Construction of plasmid pUB1

The 1.8 kb *Sst*I-*Bam*HI fragment, comprised between map coordinates 33.4-35.2 of Fig. 9, is obtained from cosmid pRP16, prepared as described in Example 18, and ligated to pUC18 previously digested with *Sst*I + *Bam*HI. The ligation mixture contains the desired plasmid pUB1.

Example 24

Construction of *E. coli* K12 XL1blue/pUB1

Approximately 10 ng of plasmid pUB1, prepared as described in Example 23, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to contain plasmid pUB1 as verified by the observation, upon agarose gel electrophoresis, of fragments 2.7 and 1.8 kb after digestion with *Sst*I + *Xba*I.

Example 25

Construction of plasmid pUC1

The 6.2 kb *Bam*HI fragment, comprised between map coordinates 54.2-60.4 kb of Fig. 9, is obtained from cosmid pRP58, prepared as described in Example 18, and ligated to pUC18 previously digested with *Bam*HI. The ligation mixture contains the desired plasmid pUC1.

Example 26

Construction of *E. coli* K12 XL1blue/pUC1

Approximately 10 ng of plasmid pUC1, prepared as described in Example 25, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that

appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUC1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 4.9 and 4.0 kb after digestion of the plasmid with *Pst*I.

Example 27

Construction of plasmid pUD1

Synthetic oligonucleotides 5'-GATCTAAGCTTGGGGG-3' and 5'-CCCCCAAGCTTA-3' are annealed and ligated to the 1.5 kb *Pst*I-*Bam*HI fragment, comprised between map coordinates 89.5-91.0 kb of Fig. 9 and obtained from cosmid pRP58, prepared as described in Example 18. The ligation mixture is digested with *Hin*dIII and ligated to pUC18 previously digested with *Pst*I + *Hin*dIII. The resulting mixture contains the desired plasmid pUD1.

Example 28

Construction of *E. coli* K12 XL1blue/pUD1

Approximately 10 ng of plasmid pUD1, prepared as described in Example 27, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to contain plasmid pUD1 as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 1.5 kb after digestion with *Pst*I + *Hin*dIII.

Example 29

Construction of plasmid pUAB1

The 0.9 kb *Eco*RI-*Sst*I fragment from plasmid pUA2, prepared as described in Example 22, and the 1.8 kb *Sst*I-*Bam*HI fragment from pUB1, prepared as described in Example 24, are ligated to pUC18 previously digested with *Eco*RI + *Bam*HI. The ligation mixture contains the desired plasmid pUAB1.

Example 30

Construction of *E. coli* K12 XL1blue/pUAB1

Approximately 10 ng of plasmid pUAB1, prepared as described in Example 29, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUAB1, as verified by the observation, upon agarose gel-electrophoresis, of two fragments of 2.7 kb after digestion of the plasmid with *Eco*RI + *Xba*I.

Example 31

Isolation of the *tetR* fragment

The 1.6 kb fragment containing the *tetR* gene is isolated after PCR amplification of pBR322 DNA using oligonucleotides 5'-GAGCTCTCATGTTTGACAGCT-3' and 5'-GAGCTCTGACTTCCGCGTTTCCAG-3' as primers, followed by digestion with *Sst*I.

Example 32

Construction of plasmid pUAB2

Plasmid pUAB1, prepared as described in Example 30, is digested with *Sst*I and ligated to the *tetR* fragment

prepared as described in Example 31. The ligation mixture contains the desired plasmid pUAB2.

Example 33

Construction of *E. coli* K12 DH5α/pUAB2

Approximately 10 ng of plasmid pUAB2, prepared as described in Example 32, are used to transform *E. coli* DH5α and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUAB2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 4.3 and 2.7 kb after digestion of the plasmid with *Eco*RI + *Xba*I.

Example 34

Construction of plasmid pUBC1

The 1.8 kb *Sst*I-*Xba*I fragment obtained from plasmid pUB1, prepared as described in Example 24, and the 4.0 kb *Xba*I-*Pst*I fragment obtained from plasmid pUC1, prepared as described in Example 26, are ligated to pUC18 previously digested with *Sst*I + *Pst*I. The ligation mixture contains the desired plasmid pUBC1.

Example 35

Construction of *E. coli* K12 XL1blue/pUBC1

Approximately 10 ng of plasmid pUBC1, prepared as described in Example 34, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUBC1, as

verified by the observation, upon agarose gel-electrophoresis, of fragments of 5.8 and 2.7 kb after digestion of the plasmid with *Eco*RI + *Hin*dIII.

Example 36

Construction of plasmid pUBC2

Plasmid pUBC1, prepared as described in Example 35 and previously digested with *Xba*I, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA polymerase and T4 DNA ligase. The ligation mixture contains the desired plasmid pUBC2.

Example 37

Construction of *E. coli* K12 DH5α/pUBC2

Approximately 10 ng of plasmid pUBC2, prepared as described in Example 36, are used to transform *E. coli* DH5α and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUBC2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 5.6 and 4.5 kb after digestion of the plasmid with *Hin*dIII.

Example 38

Construction of plasmid pUCD1

The 4.0 kb *Xba*I-*Pst*I fragment obtained from plasmid pUC1, prepared as described in Example 26, and the 1.5 kb *Pst*I-*Hin*dIII fragment isolated from plasmid pUD1, prepared as described in Example 28, are ligated to pUC18 previously digested with *Xba*I and *Hin*dIII. The

mixture contains the desired plasmid pUCD1.

Example 39

Construction of *E. coli* K12 XL1blue/pUCD1

Approximately 10 ng of plasmid pUCD1, prepared as described in Example 38, are used to transform *E. coli* XL1blue and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUCD1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 5.5 and 2.7 kb after digestion of the plasmid with *Xba*I + *Hin*dIII.

Example 40

Construction of plasmid pUCD2

Plasmid pUCD1, prepared as described in Example 39 and previously digested with *Pst*I, and the *tetR* fragment prepared as described in Example 31, are treated with T4 DNA polymerase and T4 DNA ligase. The ligation mixture contains the desired plasmid pUCD2.

Example 41

Construction of *E. coli* K12 DH5α/pUCD2

Approximately 10 ng of plasmid pUCD2, prepared as described in Example 40, are used to transform *E. coli* DH5α and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUCD2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 6.7 and 3.1 kb after

digestion of the plasmid with *Hin*dIII.

Example 42

Construction of plasmid pUAD1

The 4.3 kb *Eco*RI-*Xba*I fragment obtained from plasmid pUAB2, prepared as described in Example 33, and the 5.5 *Xba*I-*Hin*dIII fragment from plasmid pUCD1, prepared as described in Example 39, are ligated to pUC18, previously digested with *Eco*RI + *Hin*dIII. The ligation mixture contains the desired plasmid pUAD1.

Example 43

Construction of *E. coli* K12 DH5α/pUAD1

Approximately 10 ng of plasmid pUAD1, prepared as described in Example 42, are used to transform *E. coli* DH5α and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUAD1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 8.9 and 3.6 kb after digestion of the plasmid with *Hin*dIII.

Example 44

Construction of plasmid pMAB1

The 4.3 kb *Eco*RI-*Xba*I fragment obtained from plasmid pUAB2, prepared as described in Example 33, is treated with T4 DNA Polymerase and ligated to pMAK705 previously digested with *Hinc*II. The ligation mixture contains the desired plasmid pMAB1.

Example 45

Construction of *E. coli* K12 C600/pMAB1

Approximately 10 ng of plasmid pMAB1, prepared as described in Example 44, are used to transform *E. coli* C600 and a few of the resulting Cm^RTc^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pMAB1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 4.1, 3.4, 1.4 and 0.9 kb after digestion of the plasmid with *Hin*dIII + *Eco*RI.

Example 46

Construction of plasmid pMBC1

The 7.1 kb fragment from plasmid pUBC2, prepared as described in Example 37, is obtained after partial digestion with *Pst*I, treated with T4 DNA polymerase and ligated to pMAK705 previously digested with *Hinc*II. The ligation mixture contains the desired plasmid pMBC1.

Example 47

Construction of *E. coli* K12 C600/pMBC1

Approximately 10 ng of plasmid pMBC1, prepared as described in Example 46, are used to transform *E. coli* C600 and a few of the resulting Cm^RTc^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pMBC1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 9.5, 1.5, 1.3 and 0.3 kb after digestion of the plasmid with *Bam*HI.

Example 48

Construction of plasmid pMCD1

The 7.1 kb fragment from plasmid pUCD2, prepared as described in Example 41, is obtained by complete digestion with *Xba*I and partial digestion with *Hind*III, treated with T4 DNA polymerase and ligated to pMAK705, previously digested with *Hinc*II. The ligation mixture contains the desired plasmid pMCD1.

Example 49

Construction of *E. coli* K12 C600/pMCD1

Approximately 10 ng of plasmid pMCD1, prepared as described in Example 48, are used to transform *E. coli* C600 and a few of the resulting Cm^RTc^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pMCD1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 8.6 and 4.3 kb after digestion of the plasmid with *Bam*HI.

Example 50

Construction of plasmid pPAD1

The 10.0 kb *Eco*RI-*Nde*I fragment from plasmid pUAD1, prepared as described in Example 43, is ligated to pPAC-S1, prepared as described in Example 11 and previously digested with *Sca*I. The ligation mixture contains the desired plasmid pPAD1.

Example 51

Construction of *E. coli* K12 C600/pPAD1

Approximately 10 ng of plasmid pPAD1, prepared as described in Example 50, are used to transform *E. coli* C600 and a few of the resulting Km^RTc^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pPAD1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 19.7, 5.8, 3.1 and 1.2 kb after digestion of the plasmid with *Bam*HI. After digestion with *Dra*I and resolution by PFGE, pPAD1 yields fragments of 17.4, 7.4, 4.2 and 0.6 kb.

Example 52

Construction of *E. coli* K12 C600/pMAB1::pRP16

E. coli C600/pMAB1, prepared as described in Example 45, is transformed with ca. 50 ng of pRP16, prepared as described in Example 18. The Cm^RKm^R colonies that appear at 30 °C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the Cm^RKm^R colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pMAB1::pRP16, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 34, 10.7, 1.7, 1.6, 1.5, 1.2 and 0.6 kb after digestion of the plasmid with *Eco*RI.

Example 53

Construction of *E. coli* K12 C600/pMAB2

Several colonies of *E. coli* C600/pMAB1::pRP16, prepared as described in Example 52, are grown individually in LB broth containing Cm for 24 h at 30°C, diluted 1:100 and incubated for further 8 h. Appropriate dilutions are plated. Few of the resulting $Cm^RKm^STc^S$ colonies that appear at 30°C are analyzed for their plasmid content. One colony is found to carry pMAB2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 37 and 1.5 kb after digestion of the plasmid with *Eco*RI.

Example 54

Construction of *E. coli* K12 DH1/pMBC1::pRP31

Approximately 50 ng of pRP31, prepared as described in Example 18, are used to transform *E. coli* DH1 cells. Competent cells are prepared from one of the resulting Km^R colonies and transformed with ca. 10 ng of plasmid pMBC1, prepared as described in Example 47. The Cm^RKm^R colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the Cm^RKm^R colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pMBC1::pRP31, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 22.2, 14.1, 14.0 and 6.0 kb after digestion of the plasmid with *Eco*RV.

Example 55

Construction of *E. coli* K12 DH1/pMBC2

Several colonies of *E. coli* DH1/pMBC1::pRP31, prepared as described in Example 54, are grown individually in LB broth containing Cm for 24 h at 30°C, diluted 1:100 and incubated for further 8 h. Appropriate dilutions are plated. Few of the resulting $Cm^RKm^STc^S$ colonies that appear at 30°C are analyzed for their plasmid content. One colony is found to carry pMBC2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 14.4, 14.1 and 1.5 kb after digestion of the plasmid with *Eco*RI.

Example 56

Construction of *E. coli* K12 DH1/pMCD1::pRP58

Approximately 50 ng of pRP58, prepared as described in Example 18, are used to transform *E. coli* DH1 cells. Competent cells are prepared from one of the resulting Km^R colonies and transformed with ca. 10 ng of plasmid pMCD1, prepared as described in Example 48. The Cm^RKm^R colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the Cm^RKm^R colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pMCD1::pRP58, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 39, 16, 1.7, 1.6, 1.5, 1.2 and 0.6 kb

after digestion of the plasmid with *Eco*RI.

Example 57

Construction of *E.coli* K12 DH1/pMCD2

Several colonies of *E. coli* DH1/pMCD1::pRP58, prepared as described in Example 56, are grown individually in LB broth containing Cm for 24 h at 30°C, diluted 1:100 and incubated for further 8 h. Appropriate dilutions are plated. Few of the resulting $Cm^RKm^STc^S$ colonies that appear at 30°C are analyzed for their plasmid content. One colony is found to carry pMCD2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 42 and 1.5 kb after digestion of the plasmid with *Eco*RI.

Example 58

Construction of *E. coli* K12 C600/pMAB2::pPAD1

E. coli C600/pMAB2, prepared as described in Example 53, is transformed with ca. 50 ng of plasmid pPAD1, prepared as described in Example 51. The Cm^RKm^R colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the Cm^RKm^R colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pMAB2::pPAD1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 19.7, 7.2, 5.6, 5.6, 5.5, 5.2, 3.1,

2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 1.2, 0.9, 0.9 and 0.7 kb after digestion of the plasmid with *Bam*HI.

Example 59

Construction of E.coli K12 C600/pPAD2

Several colonies of *E. coli* C600/pMAB2::pPAD1, prepared as described in Example 58, are grown individually in LB containing Km for 24 h at 30°C, diluted 1:100 and incubated for further 8 h at 44°C. Appropriate dilutions are plated. Few of the resulting $Km^{R}Cm^{S}Tc^{S}$ colonies that appear at 37°C are analyzed for their plasmid content. One colony is found to carry pPAD2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 19.7, 7.2, 5.6, 5.5, 5.2, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 0.9 and 0.7 kb after digestion of the plasmid with *Bam*HI. After *Dra*I digestion and resolution by PFGE, pPAD2 yields fragments of 45, 7.4, 4.2 and 0.6 kb.

Example 60

Construction of plasmid pMCD3

The 1.4 kb *Kpn*I-*Xho*II fragment obtained from plasmid pCYPAC2 after digestion with *Xho*II, treatment with T4 DNA polymerase and digestion with *Kpn*I, and the 7.1 kb *Xba*I-*Hin*dIII fragment from pUCD2, prepared as described in Example 40 and obtained after partial digestion with *Hin*dIII, complete digestion with *Xba*I and treatment with T4 DNA polymerase, are ligated to pMAK705, previously digested with *Kpn*I + *Hinc*II. The ligation mixture contains the desired plasmid pMCD3.

Example 61

Construction of *E. coli* K12 C600/pMCD3

Approximately 10 ng of plasmid pMCD3, prepared as described in Example 60, are used to transform *E. coli* C600 and a few of the resulting Cm^RTc^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pMCD3, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 9.8 and 4.3 kb after digestion of the plasmid with *Bam*HI.

Example 62

Construction of *E. coli* K12 C600/pPAD2::pMCD3

E. coli C600/pPAD2, prepared as described in Example 59, is transformed with ca. 10 ng of plasmid pMCD3, prepared as described in Example 61. The Cm^RKm^R colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the Cm^RKm^R colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pPAD2::pMCD3, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 19.7, 9.8, 7.2, 5.6, 5.5, 5.2, 4.3, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 0.9 and 0.7 kb after digestion of the plasmid with *Bam*HI.

Example 63

Construction of *E. coli* K12 C600/pPAD3

Several colonies of *E. coli* C600/pPAD2::pMCD3, prepared as described in Example 62, are grown individually in LB broth containing Km for 24 h at 30°C, diluted 1:100 and incubated for further 8 h at 44°C. Appropriate dilutions are plated. Few of the resulting $Km^R Cm^S Tc^R$ colonies are analyzed for their plasmid content. One colony is found to carry pPAD3, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 21.5, 7.2, 5.6, 5.2, 4.3, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 0.9 and 0.7 kb after digestion of the plasmid with *Bam*HI.

Example 64

Construction of *E. coli* K12 C600/pPAD3::pMCD2

E. coli C600/pPAD3, prepared as described in Example 63, is transformed with ca. 50 ng of plasmid pMCD2, prepared as described in Example 57. The $Cm^R Km^R$ colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the $Cm^R Km^R$ colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pPAD3::pMCD2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 21.5, 10, 9,0, 7.6, 7.2, 6.2, 5.6, 5.2, 4.3, 3.1, 2.8, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 1.0,

0.9, 0.9, 0.9, 0.7, 0.5, 0.3 and 0.1 kb after digestion of the plasmid with *Bam*HI.

Example 65

Construction of *E. coli* K12 C600/pPAD4

Several colonies of *E. coli* C600/pPAD3::pMCD2, prepared as described in Example 64, are grown individually in LB broth containing Km for 24 h at 30°C, diluted 1:100 and incubated for further 8 h at 44°C. Appropriate dilutions are plated. Few of the resulting $Km^R Cm^S Tc^S$ colonies are analyzed for their plasmid content. One colony is found to carry pPAD4, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 22, 10, 7.6, 7.2, 6.2, 5.6, 5.2, 3.1, 2.8, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 1.0, 0.9, 0.9, 0.9, 0.7, 0.5, 0.3 and 0.1 kb after digestion of the plasmid with *Bam*HI. After *Dra*I digestion and resolution by PFGE, pPAD4 yields fragments of 79, 4.2 and 0.6 kb.

Example 66

Construction of *E. coli* K12 C600/pPAD4::pMBC1

E. coli C600/pPAD4, prepared as described in Example 65, is transformed with ca. 10 ng of plasmid pMBC1, prepared as described in Example 47. The $Cm^R Km^R$ colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm, aliquots are withdrawn at various times and appropriate dilutions plated. Few of the $Cm^R Km^R$ colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for

16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pPAD4::pMBC1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 22, 10, 9.6, 7.6, 7.2, 6.2, 5.6, 5.2, 3.1, 2.8, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.5, 1.4, 1.3, 1.0, 0.9, 0.9, 0.9, 0.7, 0.5, 0.3, 0.3 and 0.1 kb after digestion of the plasmid with *Bam*HI.

Example 67

Construction of *E. coli* K12 C600/pPAD5

Several colonies of *E. coli* C600/pPAD4::pMBC1, prepared as described in Example 66, are grown individually in LB broth containing Km for 24 h at 30°C, diluted 1:100 and incubated for further 8 h at 44°C. Appropriate dilutions are plated. Few of the resulting Km^R Cm^S Tc^R colonies are analyzed for their plasmid content. One colony is found to carry pPAD5, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 22, 10, 7.6, 7.2, 6.2, 5.6, 5.2, 3.1, 2.8, 2.7, 1.9, 1.9, 1.8, 1.8, 1.6, 1.4, 1.3, 1.0, 0.9, 0.9, 0.9, 0.7, 0.5, 0.3, 0.3 and 0.1 kb after digestion of the plasmid with *Bam*HI.

Example 68

Construction of *E. coli* K12 C600/pPAD5::pMBC2

E. coli C600/pPAD5, prepared as described in Example 67, is transformed with ca. 50 ng of plasmid pMBC2, prepared as described in Example 55. The Cm^RKm^R colonies that appear at 30°C on the LB-agar plates are grown at 30°C in LB broth containing Km and Cm,

aliquots are withdrawn at various times and appropriate dilutions plated. Few of the Cm^RKm^R colonies that appear on the LB-agar plates after overnight incubation at 44°C are grown in LB broth containing Km and Cm for 16 h at 44°C and analyzed for their plasmid content. One colony is found to carry pPAD5::pMBC2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 65, 33, 5.6, 4.7, 3.4, 2.8, 2.1, 1.2, 1.2, 1.0 and 0.4 kb after digestion of the plasmid with *Hin*dIII.

Example 69

Construction of *E. coli* K12 C600/pPAD6

Several colonies of *E. coli* C600/pPAD5::pMBC2, prepared as described in Example 68, are grown individually in LB broth containing Km for 24 h at 30°C, diluted 1:100 and incubated for further 8 h at 44°C. Appropriate dilutions are plated. Few of the resulting $Km^RCm^STc^S$ colonies are analyzed for their plasmid content. One colony is found to carry pPAD6, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 47, 46, 8.1, 4.6, 2.2, 0.5 and 0.1 kb after digestion of the plasmid with *Eco*RI. After digestion with *Dra*I and resolution by PFGE, pPAD6 yields fragments of 102, 4.2 and 0.6 kb.

Although the present invention is described in the Examples listed above in terms of preferred embodiments, they are not to be regarded as limiting the scope of the invention. The above Examples serve to

illustrate the principles and methodologies for assembling pre-existing overlapping segments of DNA into ESAC vectors.

It will occur to those skilled in the art that the cluster of Fig. 9 can be assembled using A-B-C-D fragments other than those specified in the Examples. Any A fragment, such that no useful genes are present to its left (using the orientation of Fig. 9) can be used for assembling the cluster. Similarly, any D fragment, such that no useful genes are present to its right (using the orientation of Fig. 9) can also be used. Furthermore, any fragment common to pRP16 and pRP31, or to pRP31 and pRP58, can be used in place of the fragments B and C, respectively, described above. It will also occur to those skilled in the art that other methods for obtaining these fragments, such as use of different segments from the cluster of Fig. 9, of different restriction endonucleases, or of the PCR, can be used for achieving equivalent results. In addition, intermediate vectors, other than the pUC-series used in the above Examples, can be used for subcloning fragments A through D, and that the use of these intermediate vectors is merely instrumental to the transfer of the fragment pairs into the *ts* vector. Some or all of the fragment pairs could therefore be cloned directly into a *ts* vector.

It will also occur to those skilled in the art that cosmids other than pRP16, pRP31 and pRP58 can be used to achieve equivalent results, provided that they encompass the entire gene cluster and they have

overlapping segments. It will also occur to those skilled in the art that pMAK705, Lorist6 and pPAC-S1, are merely examples of *ts*, cosmid and ESAC vectors, respectively. Any of the several cosmid vectors described in the literature, other *ts* replicons derived from pMAK705 or other sources, and any of the ESAC vectors other than pPAC-S1, which include the possible vectors described in Section 7.1, can be used for obtaining equivalent results.

Those skilled in the art understand that the purpose of a *ts* replicon is to select for interplasmid cointegrates at the non-permissive temperature. However, cointegrate formation can occur between any two replicons, and cointegrate can be isolated after transformation of suitable hosts with a plasmid preparation made from an *E. coli* cell harboring both replicons. Selection for the antibiotic resistance markers carried by both replicons can lead to the isolation of cointegrates from the resulting transformants.

Furthermore, it will occur to those skilled in the art that the inclusion of the *tetR* marker between the A-B, B-C and C-D fragment pairs serves solely the scope of recognizing insert exchange after resolution of the interplasmid cointegrate. Selectable markers other than *tetR* can be equally effective, as long as they are not present in the vectors. Those skilled in the art understand that the presence of a selectable marker within the fragment cassettes is not absolutely necessary, as insert exchange can be observed by other

methods, such as selective hybridization or PCR. Similarly, different *E. coli* hosts other than those used in the above Examples can be also employed.

It will also occur to those skilled in the art that, as described in Examples 58 through 69, interplasmid insert exchange can be obtained in a sequel independent of the order of the overlapping cosmid clones in the genomic contig. Indeed, the schematic of Fig. 8 illustrates the sequel of interplasmid exchanges A-B, followed by B-C and then by C-D, while Examples 58 through 69 describe the sequel A-B, C-D and last B-C. Furthermore, technical variations on the methodologies employed here can produced equivalent results. All these variations fall within the scope of the present invention.

It will occur to those skilled in the art that the principles and methodologies described in Sections 7.2.1 and 7.2.2 are not mutually exclusive. For example, a construct equivalent to pPAD6 can be directly isolated by subjecting the producer strain *P. rosea* to the principles and methodologies described in Section 7.2.1. Similarly, selected cosmids from the described *S. coelicolor* library (Redenbach et al., 1996, Mol. Microbiol. 21:77-96) can be used for assembling a large chromosomal segment into pPAC-S1, following the principles and methodologies described in Section 7.2.2. Furthermore, it will occur to those skilled in the art that the principles and methodologies of Section 7.2.1 and 7.2.2 can complement each other. For example, after constructing an ESAC

library of *P. rosea* DNA, inserts from individual ESAC clones may be enlarged by applying the principles and methodologies of Section 7.2.2, using, for example, cosmids overlapping the cognate ESAC clones.

Those skilled in the art understand that the principles and methodologies described in Section 7.2.2 and illustrated in schematic form in Fig. 4 are general enough that they can be applied to other strains and clusters. Methods for preparing high molecular weight DNA, for constructing and propagating in *E. coli* an ESAC library can be developed from the principles and methodologies described in Examples 12 through 17. Methods for preparing the appropriate combinations of fragment pairs to yield the starting plasmids described in Fig. 6, can be developed for other clusters following the principles and methodologies described in Examples 19 through 51; methods for assembling an entire cluster into an ESAC vector can be developed following the principles and methodologies described in Examples 52 through 69. In order to illustrate how the principles and methodologies described in Section 7.2 can be extended to other actinomycete strains producing different natural products, the constructions of ESAC clones carrying large gene clusters from different producer strains are reported herein. The Examples describe, for each cluster, the selection of the appropriate fragments A, B, C and D; and the construction of the starting plasmids, equivalent to those reported in Fig. 6. These plasmids can then be used to to reassemble each cluster according to the

scheme of Fig. 8.

The rapamycin gene cluster from *S. hygroscopicus* is contained within three overlapping cosmids designated cos58, cos25 and cos2 (Schwecke et al., 1995, Proc. Natl. Acad. Sci. USA 92:7839-7843). The Examples described herein report the preparation of the appropriate fragments A, B, C and D; the construction of the plasmids containing the A-B, B-C and C-D cassettes; and the cloning approach to obtain constrcuts equivalent to those reported in Fig. 6.

Example 70

Preparation of the rapamycin fragments A, B, C and D

Primers 5'-TTTTTGAATTCGGTACCAGCCGACGGCGA-3' and 5'-TTTTTGGATCCCTGTTCCACCAGCGCACC-3' are used to amplify a 1.2 kb fragment from cos58; primers 5'-TTTTTGGATCCAGGAAGCCCTGTGCTGTC-3' and 5'-TTTTTTCTAGACCG-TCGTCGGTGGTTCTC-3' a 1.2 kb fragment from cos58; primers 5'-TTTTTTGTAGAGGTCAAGATCCGGGGCAT-3' and 5'-TTTTTCTGCAGGACAGCGCCCTTGAGGTG-3' a 1.2 kb fragment from cos25; and primers 5'-TTTTTCTGCAGGCGACGAAGAGGGGC-3' and 5'-TTTTTAAGCTTAGCGCGACCGGGGCGGT-3' a 0.9 kb fragment from cos2. Fragment A, B, C and D are then digested with *Eco*RI + *Bam*HI, *Bam*HI + *Xba*I, *Xba*I + *Pst*I, and *Pst*I + *Hin*dIII, respectively.

Example 71

Construction of plasmid pUR1

Fragments A and B, prepared as described in Example 70, are ligated to pUC18 digested with *Eco*RI + *Xba*I. The resulting mixture contains the desired plasmid pUR1.

Example 72

Construction of *E. coli* K12 DH1/pUR1

Approximately 10 ng of plasmid pUR1, prepared as described in Example 71, are used to transform *E. coli* DH1 and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 2.4 kb after digestion of the plasmid with *Eco*RI + *Xba*I.

Example 73

Construction of plasmid pUR2

Fragments B and C, prepared as described in Example 70, are ligated to pUC18 digested with *Bam*HI + *Pst*I. The resulting mixture contains the desired plasmid pUR2.

Example 74

Construction of *E. coli* K12 DH1/pUR2

Approximately 10 ng of plasmid pUR2, prepared as described in Example 73, are used to transform *E. coli* DH1 and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 2.4 kb after digestion of the plasmid with *Bam*HI + *Pst*I.

Example 75

Construction of plasmid pUR3

Fragments C and D, prepared as described in Example 70, are ligated to pUC18 digested with *Xba*I + *Hind*III. The resulting mixture contains the desired plasmid pUR3.

Example 76

Construction of *E. coli* K12 DH1/pUR3

Approximately 10 ng of plasmid pUR3, prepared as described in Example 75, are used to transform *E. coli* DH1 and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR3, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 2.1 kb after digestion of the plasmid with *Eco*RI + *Hind*III.

Example 77

Construction of plasmid pUR11

Plasmid pUR1, prepared as described in Example 72 and previously digested with *Bam*HI, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA Polymerase and DNA ligase. The ligation mixture contains the desired plasmid pUR11.

Example 78

Construction of *E. coli* K12 DH1/pUR11

Approximately 10 ng of plasmid pUR11, prepared as described in Example 77, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their

plasmid content. One colony is found to carry pUR11, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 3.9 and 2.8 kb after digestion of the plasmid with *Hin*dIII.

Example 79

Construction of the plasmid pUR21

Plasmid pUR2, prepared as described in Example 74 and previously digested with *Xba*I, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA Polymerase and DNA ligase. The ligation mixture contains the desired plasmid pUR21.

Example 80

Construction of *E. coli* K12 DH1/pUR21

Approximately 10 ng of plasmid pUR21, prepared as described in Example 79, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR21, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 3.9 and 2.8 kb after digestion of the plasmid with *Hin*dIII.

Example 81

Construction of the plasmid pUR31

Plasmid pUR3, prepared as described in Example 76 and digested with *Pst*I, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA Polymerase and DNA ligase. The ligation mixture

contains the desired plasmid pUR31.

Example 82

Construction of *E. coli* K12 DH1/pUR31

Approximately 10 ng of plasmid pUR31, prepared as described in Example 81, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR31, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 3.9 and 2.5 kb after digestion of the plasmid with *Hin*dIII.

Example 83

Construction of plasmid pUR13

The 4.0 kb *Eco*RI-*Xba*I fragment obtained from plasmid pUR11, prepared as described in Example 78, and the 2.1 kb *Xba*I-*Hin*dIII fragment obtained from plasmid pUR3, prepared as described in Example 76, are ligated to pUC18 digested with *Eco*RI + *Hin*dIII. The ligation mixture contains the desired plasmid pUR13.

Example 84

Construction of *E. coli* K12 DH1/pUR13

Approximately 10 ng of plasmid pUR13, prepared as described in Example 83, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR13, as verified by the observation, upon agarose gel-

electrophoresis, of fragments of 4.9 and 3.9 kb after digestion of the plasmid with *Hin*dIII.

Those of ordinary skill in the art understand that the plasmids constructed above can be used for transfering the two-fragment cassettes present in pUR11, pUR21 and pUR31 into a *ts* vector. This can be achieved by recovering the 4.0 kb insert from pUR11, the 4.0 kb insert from pUR21, and the 3.7 kb insert from pUR31, after digestion with *Eco*RI + *Xba*I, *Eco*RI + *Pst*I, and *Eco*RI + *Nde*I, respectively. Similarly, those of ordinary skill in the art understand that the 6.1 kb four-fragment cassette present in plasmid pUR13 can be easily transfered into pPAC-S1 after digestion with *Eco*RI + *Nde*I. These subcloning experiments lead to the formation of plasmids equivalent to those reported in Fig. 6.

As another application of the principles and methodologies of the present invention, the Examples reported below describe the preparation of the appropriate fragments A, B, C and D from the *Sac. erythraea* erythromycin gene cluster. This cluster has been described and is contained within a series of overlapping clones (Tuan et al., 1990, Gene 90:21-29; Donadio et al., 1993, In *Industrial Microorganisms: Basics and Applied* Genetics, Baltz, Hegeman and Skatrud eds., ASM, Washington, DC, pp.257-265; Pereda et al., 1997, Gene 193:65-71). The construction of the plasmids containing the A-B, B-C and C-D cassettes and the cloning approach to obtain constructs equivalent to

those reported in Fig. 6 are also described.

Example 85

Preparation of the erythromycin fragment A

Synthetic oligonucleotides 5'-CATGGGAATTCGGGGG-3' and 5'-CCCCCGAATTCC-3' are annealed and ligated to the 1.2 kb *Nco*I-*Bam*HI fragment isolated from cosmid p3B2. The resulting mixture is digested with *Eco*RI + *Bam*HI.

Example 86

Preparation of the erythromycin fragments B, C and D

Primers 5'-TTTTTGGATCCGGGGCAGCGGTTGGTTCC-3' and 5'-TTTTTTCTAGAAGGCAGCTCCAGATGATC-3' are used to amplify a 1.0 kb fragment from cosmid p3B2; primers 5'-TTTTTCTAGACCGGACTCGGCCGGCTCG-3' and 5'-TTTTTCTGCAGCCGCACGCCTCGGTGGTC-3' a 1.1 kb fragment from cosmid pS1; and primers 5'-TTTTTCTGCAGGGACCCTGAGTGCAGGTG-3' and 5'-TTTTTAAGCTTCAGTAGCCGTCGCTGAGC-3' a 1.1 kb fragment from plasmid pEB6. Fragments B, C and D are then digested with *Bam*HI + *Xba*I, *Xba*I + *Pst*I, and *Pst*I + *Hin*dIII, respectively.

Example 87

Construction of plasmid pUE1

Fragment A, prepared as described in Example 85, and fragment B, prepared as described in Example 86, are ligated to pUC18 digested with *Eco*RI + *Xba*I. The resulting mixture contains the desired plasmid pUE1.

Example 88

Construction of *E. coli* K12 DH1/pUE1

Approximately 10 ng of plasmid pUE1, prepared as described in Example 87, are used to transform *E. coli* DH1 and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUR1, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 2.2 kb after digestion of the plasmid with *Eco*RI + *Xba*I.

Example 89

Construction of plasmid pUE2

Fragments B and C, prepared as described in Example 86, are ligated to pUC18 digested with *Bam*HI + *Pst*I. The resulting mixture contains the desired plasmid pUE2.

Example 90

Construction of *E. coli* K12 DH1/pUE2

Approximately 10 ng of plasmid pUE2, prepared as described in Example 89, are used to transform *E. coli* DH1 and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUE2, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 2.1 kb after digestion of the plasmid with *Bam*HI + *Pst*I.

Example 91

Construction of plasmid pUE3

Fragments C and D, prepared as described in Example 86, are ligated to pUC18 digested with *Xba*I + *Hind*III. The

resulting mixture contains the desired plasmid pUE3.

Example 92

Construction of *E. coli* K12 DH1/pUE3

Approximately 10 ng of plasmid pUE3, prepared as described in Example 91, are used to transform *E. coli* DH1 and a few of the resulting Ap^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUE3, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 2.7 and 2.2 kb after digestion of the plasmid with *Eco*RI + *Hin*dIII.

Example 93

Construction of plasmid pUE11

Plasmid pUE1, prepared as described in Example 88 and previously digested with *Bam*HI, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA Polymerase and DNA ligase. The ligation mixture contains the desired plasmid pUE11.

Example 94

Construction of *E. coli* K12 DH1/pUE11

Approximately 10 ng of plasmid pUE11, prepared as described in Example 93, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUE11, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 3.9 and 2.6 kb after

digestion of the plasmid with *Hin*dIII.

Example 95

Construction of the plasmid pUE21

Plasmid pUE2, prepared as described in Example 90 and previously digested with *Xba*I, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA Polymerase and DNA ligase. The ligation mixture contains the desired plasmid pUE21.

Example 96

Construction of *E. coli* K12 DH1/pUE21

Approximately 10 ng of plasmid pUE21, prepared as described in Example 95, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUE21, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 3.7 and 2.7 kb after digestion of the plasmid with *Hin*dIII.

Example 97

Construction of the plasmid pUE31

Plasmid pUE3, prepared as described in Example 92 and digested with *Pst*I, and the *tetR* fragment, prepared as described in Example 31, are treated with T4 DNA Polymerase and DNA ligase. The ligation mixture contains the desired plasmid pUE31.

Example 98

Construction of *E. coli* K12 DH1/pUE31

Approximately 10 ng of plasmid pUE31, prepared as described in Example 97, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUE31, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 3.8 and 2.7 kb after digestion of the plasmid with *Hin*dIII.

Example 99

Construction of plasmid pUE13

The 3.8 kb *Eco*RI-*Xba*I fragment obtained from plasmid pUE11, prepared as described in Example 94, and the 2.2 kb *Xba*I-*Hin*dIII fragment obtained from plasmid pUE3, prepared as described in Example 92, are ligated to pUC18 digested with *Eco*RI + *Hin*dIII. The ligation mixture contains the desired plasmid pUE13.

Example 100

Construction of *E. coli* K12 DH1/pUE13

Approximately 10 ng of plasmid pUE13, prepared as described in Example 99, are used to transform *E. coli* DH1 and a few of the resulting Tc^RAp^R colonies that appear on the LB-agar plates are analyzed for their plasmid content. One colony is found to carry pUE13, as verified by the observation, upon agarose gel-electrophoresis, of fragments of 4.8 and 3.9 kb after digestion of the plasmid with *Hin*dIII.

Those of ordinary skill in the art understand that the plasmids constructed above can be used for transfering the two-fragment cassettes present in pUE11, pUE21 and pUE31 into a *ts* vector. This can be achieved by recovering the 3.8 kb insert from pUE11, the 3.7 kb insert from pUE21, and the 3.8 kb insert from pUE31, after digestion with *Eco*RI + *Xba*I, *Eco*RI + *Pst*I, and *Eco*RI + *Nde*I, respectively. Similarly, those of ordinary skill in the art understand that the 6.0 kb four-fragment cassette present in plasmid pUE13 can be easily transfered into pPAC-S1 after digestion with *Eco*RI + *Nde*I. These subcloning experiments lead to the formation of plasmids equivalent to those reported in Fig. 6.

The Examples reported above describe the principle and methodologies for assembling the erythromycin gene cluster into the ESAC vector pPAC-S1. It will occur to those skilled in the art that the principles and methodologies illustrated in Fig. 7 and described in Examples 52 through 69 can be applied to the erythromycin gene cluster, employing the pMAK705 derivatives constructed according to the principles described above and the eryhtomycin cosmids.

As a further example, the preparation of the appropriate fragments A, B, C and D from the *A. mediterranei* rifamycin gene cluster is illustrated below. This cluster has been described and is contained within a series of overlapping clones (August et al., 1998, Chem. Biol. 5:69-79).

Example 101

Preparation of the rifamycin fragments A, B, C and D

Primers 5'-TTTTTGAATTCTGCAGACCGCCGAGGAAG-3' and 5'-TTTTTGGATCCGGAGTCGTAGCTGACGAC-3'; 5'-TTTTGGATCCCGACCACGCGGGACGTC-3' and 5'-TTTTTTCTAGACCAGGGAACCCGTGCTGC-3'; 5'-TTTTTTCTAGACGGAAGCTCGCCGCGATC-3' and 5'-TTTTTCTGCAGGTCCGTAGCCCGGACACC-3'; and 5'-TTTTTCTGCAGTTCGGGCGACAGTTCCTT-3' and 5'-TTTTTAAGCTTCAACAAGCCATCCGGGTC-3', are used to amplify fragments of 1.2, 1.2, 1.2 and 1.0 kb, respectively, from *A. mediterranei* genomic DNA. Fragments A, B, C and D are then digested with *Eco*RI + *Bam*HI, *Bam*HI + *Xba*I, *Xba*I + *Pst*I, and *Pst*I + *Hin*dIII, respectively.

Those of ordinary skills in the art understand that the fragments generated from the rifamycin gene cluster contain the same restriction sites as those generated from the rapamycin and erythromycin gene clusters, so that the same cloning strategies for generating the pUC18 derivatives containing the A-B, B-C and C-D cassettes, described above in Examples 72-77 for the rapamycin cluster and 88-93 for the erythromycin cluster, can also be applied to the rifamycin cluster. In addition, the rifamycin fragments A, B, C and D have been selected so that the same cloning methodologies described above for inserting *tetR* within the A-B, B-C and C-D cassettes from the rapamycin and erythromycin clusters, described in Examples 78-83 and 94-99, respectively, can be applied in this instance as well. Furthermore, the construction

of the four-fragment cassette can also make use of the same cloning strategy. Therefore, following the same principles and methodologies described in detail for the rapamycin and erythromycin clusters in Examples 72-85 and 88-101, respectively, plasmids equivalent to those reported in Fig. 6 can be constructed for assembling the rifamycin cluster into the ESAC vector pPAC-S1. It will occur to those skilled in the art that the principles and methodologies illustrated in Fig. 7 and described in Examples 52-69 for the *P. rosea* cluster can be applied to any gene cluster, once the appropriate pMAK705 derivatives have been constructed, employing available overlapping clones.

Thus, as schematized in Fig. 7, interplasmid insert exchange can be conducted between any plasmid containing the desired region and the cognate *ts* construct. Plasmids corresponding to pAB2, pBC2 and pCD2 can therefore be derived from any cluster. Similarly, the principles and methodologies illustrated in Fig. 8 can be applied employing the appropriate A-B-C-D cassette and the cognate pMAK705 derivatives prepared according to the scheme of Fig. 7. The principles and methodologies illustrated in Fig. 7 and Fig. 8 and described in Examples 52-69 can therefore be extended to other clusters.

It will occur to those skilled in the art that, although illustrated in Fig. 5 through 8 by three overlapping clones and described in the Examples 58 through 69 by the use of five rounds of interplasmid insert exchange, the principles and methodologies

described in this section of the present invention can be extended to a different number of overlapping clones. If *n* is the number of overlapping clones that encompass the desired genomic segment, *n* will also be the number of homologous recombination rounds that introduce cluster DNA into the ESAC vector. If an Ab^R marker is used to facilitate monitoring insert exchange, the total number of rounds of homologous recombination will be equal to *2n - 1*. Interplasmid homologous recombination has been described to introduce large DNA segments into a desired vector (O'Connor et al., 1989, Science 244:1307-1312; Kao et al., 1994, Science 265:509-512) or to target a smaller segment into a large episome (Yang et a., 1997, Nature Biotechnol. 15:859-865). However, it was not be anticipated that these procedures could be applied iteratively for the precise reconstruction of very large DNA segments.

7.3 Identification of positive clones

The principles and methodologies described in Section 7.2 for obtaining an entire gene cluster in an ESAC vector rely on the identification of the desired genomic segment. When using the principles and methodologies described in Section 7.2.1, the desired clones are identified by screening an ESAC library with one of the possible strategies described below. When using the principles and methodologies described in Section 7.2.2, the desired clones are identified in a genomic library, such as a cosmid library, with one of

the possible strategies described below, and then assembled into ESAC. The principles and methodologies for identifying the genes responsible for the biosynthesis of natural products are well described in the literature and are reported here solely to illustrate the fact that they represent a necessary step in the overall scope of the present invention.

The genes involved in the biosynthesis of natural products in actinomycetes are invariably found as gene clusters in the chromosome of the producing organism, often associated with one or more resistance determinants. Consequently, identifying one gene allows ready access to all the others. One or more genes responsible for the biosynthesis of a natural product could have been described, or the entire cluster could be known. Several biosynthesis clusters from actinomycetes have been reported and other clusters are likely to be described in the future. Suitable probes from the cluster extremities can be derived from published clusters, when available. Thus, fragments A and D, described in Example 70, can be used as probes to screen an ESAC library prepared from *S. hygroscopicus* DNA. ESAC clones positive to both probes will contain the rapamycin cluster. Similar strategies can be applied to ESAC libraries prepared from *Sac. erythraea* and *A. mediterranei* DNA, screened with fragments A and D, prepared as described in Examples 85-86 and 101, respectively.

If no biosynthesis genes are known, different strategies for identifying them can be applied. These

strategies are well described in the literature and are summarized below. One possible strategy involves the isolation of the resistance gene(s) after cloning in a heterologous host that is sensitive to that natural product (for example, Stanzak et al., 1986, Bio/Technol. 4:229-232). Another possible strategy is based on reverse genetics: a particular biosynthetic enzyme is purified, and from its partial protein sequence(s) the corresponding gene is isolated via PCR or hybridization (for example, Fishman et al., 1987, Proc. Natl. Acad. Sci. USA 84:8248-8252). Another approach relies on the complementation of mutants blocked in one or more biosynthesis steps, after introduction of a DNA library constructed in a suitable vector into the wild type strain (for example, Malpartida and Hopwood, 1984, Nature 309:462-464). Another approach involves the construction of an expression library in a suitable vector in an appropriate host, where the gene product is sought after using specific antibodies or looking for a particular enzymatic activity (for example, Jones and Hopwood, 1984, J. Biol. Chem. 259:14151-14157). Another possible approach makes use of heterologous probes derived from biosynthesis, resistance or regulatory genes. Natural products can be broadly grouped into classes according to their biosynthetic origin, and for many of them suitable probes are available. For example, genes encoding aromatic or modular polyketide synthases can be effectively identified through the use of heterologous hybridization probes (Malpartida et

al., 1987, Nature 325:818-821; Schwecke et al., 1995, Proc. Natl. Acad. Sci. USA 92:7839-7843); suitable probes have been reported for peptide synthetase genes (Turgay and Marahiel, 1994, Pept. Res. 7:238-241); for genes involved in the formation or attachment of modified sugars (Decker et al., 1996, FEMS Microbiol. Lett. 141:195-201). As the understanding of the genetics of natural product biosynthesis increases, other heterologous probes will become available.

The size of clusters can be estimated from those of known clusters involved in the synthesis of structurally similar natural products. For examples, synthesis of macrolides is expected to require clusters in the 60-70 kb range (Katz and Donadio, 1993, Annu. Rev. Microbiol. 47:875-912; Kuhstoss et al., 1996, Gene 183:231-236); synthesis of glycopeptides, clusters in the 70 kb range (van Wageningen et al., 1998, Chem. Biol. 5:155-162). In instances where no clusters have been described for the same structural class of natural products, the size of the relevant cluster can be estimated from considerations about its known or likely biosynthesis route. Once the desired cluster has been identified, its extent can be established by analysis of the DNA sequence of the cloned cluster or of parts thereof. Comparison of the DNA sequence to databases can allow the identification of the likely borders of the gene cluster.

Employing the above mentioned approaches, the desired gene cluster can be identified in any library. If an ESAC library is used, the identified cluster is

ready for transfer into the production host. If a smaller fragment library is employed, the cluster can be assembled into an ESAC vector.

Those skilled in the art understand that, when an ESAC library from a donor organism is constructed, any ESAC clone can be selected from said library and transferred into a production host. Therefore, a single donor organism can be utilized as the source of several biosynthesis clusters that can be mobilized into a production host. Similarly, an ESAC library needs not be constructed from a single donor organism.

7.4 Transformation of a *Streptomyces* host

Once the desired gene cluster has been introduced into an ESAC vector, one or more ESAC clones are introduced into a suitable *Streptomyces* host. This is accomplished by employing published procedures for transformation of *Streptomyces*. Only minor modifications from established procedures (Hopwood et al., 1985, *Genetic Manipulation of Streptomyces: A Laboratory Manual*, The John Innes Foundation, Norwich, UK) are required for obtaining a sufficient number of transformants. Because transformations are performed with single, purified ESAC clones, transformation efficiencies do not need to be particularly high. The Examples reported below illustrate the principles and methodologies for introducing ESAC clones into *S. lividans*. They serve to describe the present invention and are not meant to restrict its scope. *Streptomyces* transformants are selected for Th^R, specified by the *tsr* marker present

in the ESAC vector. Since the incoming DNA is incapable of self-replication in *Streptomyces*, site-specific integration occurs at the chromosomal *attB* site, mediated by the *int-attP* function specified by the ESAC vector. That integration has occurred at the proper site can be verified by Southern hybridization or by PFGE analysis of the transformants. Fig. 11 illustrates a PFGE separation of a *S. lividans* derivative carrying an ESAC clone with a 70 kb insert integrated into its chromosome.

Example 102

Introduction of ESAC clones into *S. lividans* ZX7

A few hundred ng of three individual ESAC clones, prepared as described in Example 17 and carrying inserts of *S. coelicolor* DNA of 70, 120 and 140 kb (designated ESAC-70, ESAC-120, and ESAC-140, respectively), are used to transform protoplasts of *S. lividans* ZX7. The colonies that appear on the R2YE plates, after overlaying with Th, are analyzed for their Th^R by streaking them on fresh R2YE plates.

Example 103

Cultivation and preservation of *S. lividans* ZX7/ESAC

Individual colonies of *S. lividans* ZX7 transformants with the individual ESAC clones, prepared as described in Example 102, are grown for several passages in solid medium without and with Th. Spore suspension, or mycelium prepared after cultivation in JM or YEME medium with Th, are stored at -80°C after addition of

glycerol to 20% (v/v).

Example 104

Characterization of *S. lividans* ZX7 *attB*::ESAC-70

Individual colonies of *S. lividans* ZX7 *attB*::ESAC-70, prepared as described in Example 102, are grown in YEME and total genomic DNA is prepared. The DNA is digested with *Bam*HI, resolved by agarose gel-electrophoresis, and transferred onto a membrane. Hybridization to labeled pPAC-S1 DNA, prepared as described in Example 11, reveals the appearance of three bands of approximately 16, 8 and 2.7 kb. PFGE analysis of genomic DNA reveals the disappearance of a 2.5 Mb *Dra*I fragment present in ZX7 and the appearance of two fragments of 1.4 and 1.1 Mb (Fig. 11).

Although the present invention is described in the Examples listed above in terms of preferred embodiments, they are not to be regarded as limiting the scope of the invention. The above Examples serve to illustrate the principles and methodologies for introducing ESAC clones into *S. lividans*, for cultivating the resulting transformants and for confirming their genotype. The above Examples serve to illustrate the principles and methodologies for the transformation of *S. lividans* with ESAC clones carrying DNA inserts from a different species. It will occur to those skilled in the art that additional ESAC clones, either containing different inserts of *S. coelicolor* DNA, prepared as described in Example 17, or carrying

DNA inserts from other actinomycetes can be used to transform *S. lividans* ZX7. As another example transfer of large DNA segments, the transformation of *S. lividans* with a *P. rosea* gene cluster is illustrated below. Confirmation of the correct genotype of the resulting transformants is illustrated in Fig. 12.

Example 105

Construction *S. lividans* ZX7 *attB*::pPAD6

A few hundred ng of pPAD6, prepared as described in Example 69, are used to transform protoplasts of *S. lividans* ZX7. The colonies that appear on the R2YE plates, after overlaying with Th, are analyzed for their Th^R by streaking them on fresh R2YE plates.

Example 106

Characterization of *S. lividans* ZX7 *attB*::pPAD6

Individual colonies of *S. lividans* ZX7 *attB*::pPAD6, prepared as described in Example 105, are grown in YEME medium and total DNA is prepared. The DNA is digested with *Bam*HI, resolved by agarose gel-electrophoresis and transferred onto a membrane. Hybridization to labeled pPAD6 il illustrated in Fig. 12. Bands of 22, 10, 7.6, 7.2, 6.2, 5.6, 5.2, 3.1, 3.0, 2.8, 2.7, 2.6, 2.5, 2.1, 1.9, 1.9, 1.8, 1.6, 1.5, 1.4, 1.2, 1.0, 1.0, 0.9, 0.9, 0.9, 0.7, 0.6, 0.5, 0.5, 0.3 and 0.1 kb. The profile of *P. rosea* DNA is shown for comparison.

Those skilled in the art understand that *S. lividans* ZX7 *attB*::pPAD6 contains the expected number

and size of bands expected from transfer of the cluster of Fig. 9 via pPAD6. In analogy to the above Examples, the rapamycin, erythromycin and rifamycin clusters assembled in ESAC, according to the principles and methodologies described in Section 7.2.2, can be used to transform *S. lividans*. It will occur to those skilled in the art that other *S. lividans* strains can be equally used as hosts for transformation with ESAC clones. Furthermore, ΦC31 can lysogenize other *Streptomyces* species, in addition to *S. lividans*. These include but are not limited to the species reported in Table 2. Furthermore, a ΦC31 *attB* site may be engineered into *Streptomyces* species or other actinomycetes that are not naturally lysogenized by phage ΦC31. Therefore, any ESAC clone, prepared according to the principles and methodologies of Section 7.2, and any natural or engineered actinomycete host, fall within the scope of the present invention.

It will occur to those skilled in the art that alternative methods for introducing DNA into an actinomycete host can be employed. These include but are not limited to electroporation (MacNeil, 1989, FEMS Microbiol. Lett. 42:239-244) and conjugation from *E. coli* (Mazodier et al., 1989, J. Bacteriol. 171:3583-3585). It will also occur to those skilled in the art that alternative media and growth conditions can be employed for cultivating the transformants, and that they can be analyzed by different methods than those described above. Technical variations on the methodologies described above can produced equivalent

results. All these variations fall within the scope of the present invention.

Table 2

List of exemplary species of *Streptomyces* and other genera of *Actinomycetales* allowing *attP*-mediated integration of ΦC31 (Hopwood et al., 1985, Genetic Manipulation of Streptomyces: A Laboratory Manual, The John Innes Foundation, Norwich, UK; Lomovskaya et al., 1997, Microbiol. 143:875-883; Kuhstoss et al., 1991, Gene 97:143-146; Soldatova et al., 1994, Antibiot. Khimioter. 39:3-7).
Streptomyces coelicolor
Streptomyces lividans
Streptomyces hygroscopicus
Streptomyces bambergiensis
Streptomyces ambofaciens
Streptomyces griseofuscus
Streptomyces lipmanii
Streptomyces thermotolerans
Streptomyces clavuligerus
Streptomyces fradiae
Saccharopolyspora spinosa

7.5 Growth of the recombinant *Streptomyces* and metabolite production

When an ESAC clone, introduced into a production host according to the principles and methodologies described in Section 7.4, carries the entire biosynthesis gene cluster derived from a donor organism, the recombinant strain produces the relevant natural product. Naive actinomycete hosts have been shown to produce the appropriate natural product or its intermediate(s) when the relevant DNA was introduced into them (Malpartida and Hopwood, 1984, Nature 309:462-464; Hong et al., 1997, J. Bacteriol. 179:470-476; Kao et al., 1994, Science 265:509-512). Thus, transformants of *Streptomyces* and other actinomycete species carrying the relevant biosynthesis clusters are expected to produce the corresponding natural product. The recombinant production hosts are cultivated in a suitable medium and the presence of the relevant metabolites is determined following appropriate procedures, which may include biological assays, chromatographic properties, MS, NMR, etc.

It will occur to those skilled in the art that ESAC clones, containing the relevant biosynthesis cluster derived from any donor actynomycete, can be used to transform *S. lividans*. The resulting transformants will produce the corresponding natural product. For example, an ESAC clone carrying the rapamycin, erythromycin or rifamycin cluster, prepared according to the principles of Section 7.2, can be used to transform *S. lividans* and rapamycin, erythromycin or

rifamycin, respectively, can be produced by the resulting recombinant strain. Furthermore, it will occur to those skilled in the art that other *Streptomyces* or actinomycete strains that naturally contain or have been engineered to contain a phage ΦC31 *attB* site, can be used as production hosts for desired natural products. Therefore, any natural product produced after introduction of the relevant cluster carried on ESAC into a suitable production host, falls within the scope of the present invention.

The present invention describes principles and methodologies for optimizing and speeding up the process of lead optimization and development in drug discovery. These can be applied since the early stages of drug discovery as briefly summarized herein. A natural product produced by a donor organism has an interesting property, such as antibacterial, antifungal, antitumor, antihelmintic, herbicidal, immunosuppressive, or other pharmacological activity. The potential is seen for increasing the productivity of the producing organism, and/or for improving the biological or physico-chemical properties of said natural product after manipulating its structure. The biosynthetic pathway for the natural product is inferred from its chemical structure. This leads to a hypothesis on the genes involved, including the approximate size of the corresponding cluster. A large insert library is constructed in the ESAC vectors described herein using genomic DNA prepared from the donor organism. Through a judicious choice of

hybridization probes and PCR primers, the desired cluster is identified in the library. Alternatively, the cluster is assembled into the ESAC vectors described herein from overlapping cosmid clones identified by hybridization as above. The selected clone(s) are transferred into *S. lividans*, *S. coelicolor* or other suitable strain, and the resulting transformants are evaluated for production of the natural product.

Once production is obtained, the desired genetic, physiological and technological manipulations can be performed on the production host, employing well-developed methodologies. The bioactive molecule is purified from a known host, amid a background of known metabolites. If necessary, *ad hoc* mutations can be conveniently introduced in the production host to eliminate unwanted, interfering products. Because of the deeper knowledge on the physiological processes and regulatory networks for secondary metabolism in the production host compared to the donor organism, targeted approaches to strain improvement, using classical and molecular techniques, can be applied. Furthermore, well-characterized mutant strains are available for the production host, and the desired ESAC clone could be easily introduced into different genetic backgrounds. In addition, the biosynthetic pathway can be easily manipulated, because of the availability of the cloned genes and the efficient genetic tools for the production host. Finally, additional specialized genes or even entire clusters can be introduced into

the production host, further expanding the possible applications of the present invention.

Finally, even in a case where the natural product may not be made by the production host after transfer of the relevant cluster, appropriate tools are available to remedy that situation. Lack of production of the expected natural product might be due to several possibilities: absence of required gene(s); DNA, gene product or natural product instability; inadequate levels of gene expression or of appropriate precursors. In a well-defined production host, each of these possible causes may be directly investigated and remedied.

Therefore, the present invention provides significant advantages over the existing process of drug discovery and development, including production. It exploits the fact that the host where the natural product will be produced is an organism commonly used for process development and genetic manipulation, with substantial information available, including safety for handling it.

SEQUENCE LISTING

<110> BIOSEARCH ITALIA SpA

<120> Methods for Transferring the Capability to Produce a Natural Product into a Suitable Production Host

<130> G67713 RS/mg

<140>
<141>

<160> 2

<170> PatentIn Ver. 2.1

<210> 1
<211> 200
<212> DNA
<213> Planobispora rosea

<400> 1
ggatcccgag caccgaccag ccgtgggcgg ggacgagaca cgggtctccc ggagcctccc 60
ccgacgactc cagcacggcc aggcccgcgg cctcgaccgg gaagcggtag ggcctgtcgt 120
ccacggttga gcagggtgag cagtgcccgg ccgggatggt ccgggtcagc cgaggccagc 180
gcggcggccc ggttgctcag 200

<210> 2
<211> 200
<212> DNA
<213> Planobispora rosea

<400> 2
ccgggagatc cgccgacgcc ggcggccgtg caccacggtc ctcctgttcc gggcgacggt 60
gaacggcagg caggtccacg gttccgattt cctgcacttc gacgacgacg gcctcatcgg 120
cgagctcacc gtcatggtcc ggccgctgtc gggagcgaac gcgctggccg aggcgatggg 180
cgcccagttc gaacggatcc 200

8. CLAIMS

We claim:

1. A method for transferring the formation of a natural product from an actinomycete donor organism that is the original producer of said natural product to a different actinomycete production host, where this transfer is achieved through the use of an *E. coli-Streptomyces* artificial chromosome.

2. An actinomycete strain that is constructed from a production host after transfer of a cluster from a donor organism as recited in claim 1.

3. An actinomycete strain as recited in claim 2 that is *Streptomyces lividans*.

4. A gene cluster governing the biosynthesis of a natural product from a donor organism as recited in claim 1 that is carried on an *E. coli-Streptomyces* artificial chromosome.

5. A modified strain obtained after manipulation of the production host as recited in claim 2.

6. A modified production host as recited in claim 5 that is *Streptomyces lividans*.

7. An *E. coli* host that contains an *E. coli-Streptomyces* artificial chromosome carrying a gene cluster governing the biosynthesis of a natural product as recited in claim 3 and that can be used as a source of DNA for transformation of a production host as recited in claim 2.

8. An *E. coli-Streptomyces* artificial chromosome that carries a gene cluster governing the biosynthesis

of a natural product as recited in claim 3 and that can be used for the purpose recited in claim 1.

9. A library constructed in an *E. coli-Streptomyces* artificial chromosome as recited in claim 5 using DNA prepared from an actinomycete donor organism, or a mixture of donor organisms.

10. An *E. coli-Streptomyces* artificial chromosomes as recited in claim 5 that is the vector pPAC-S1 described in Fig. 2.

11. An *E. coli-Streptomyces* artificial chromosomes as recited in claim 5 that is the vector pPAC-S2 described in Fig. 2.

12. An actinomycete strain as recited in claim 2 that carries the gene cluster reported in Fig. 9.

13. An actinomycete strain as recited in claim 12 that is *Streptomyces lividans*.

14. An actinomycete strain as recited in claim 12 that carries the construct pPAD6.

15. An actinomycete strain as recited in claim 14 that is *Streptomyces lividans*.

16. A modified strain as recited in claim 5 that is obtained after manipulation of the production host of claim 14.

17. A modified strain as recited in claim 16 that is *Streptomyces lividans*.

18. An *E. coli-Streptomyces* artificial chromosome as recited in claim 3 that carries the gene cluster reported in Fig. 9.

19. An *E. coli-Streptomyces* artificial chromosome as recited in claim 16 that is the construct pPAD6.

20. An actinomycete production host as recited in claim 2 that contains the *E. coli-Streptomyces* artificial chromosome carrying the rapamycin gene cluster.

21. An actinomycete production host as recited in claim 20 that is *Streptomyces lividans*.

22. A modified strain as recited in claim 5 obtained after manipulation of the production host as recited in claim 20.

23. A modified strain as recited in claim 22 that is *Streptomyces lividans*.

24. An *E. coli-Streptomyces* artificial chromosome as recited in claim 3 that carries the rapamycin gene cluster.

25. An *E. coli-Streptomyces* artificial chromosome as recited in claim 24 that contains the vectors pPAC-S1 or pPAC-S2.

26. An actinomycete production host as recited in claim 2 that contains the *E. coli-Streptomyces* artificial chromosome carrying the erythromycin gene cluster.

27. An actinomycete production host as recited in claim 26 that is *Streptomyces lividans*.

28. A modified strain as recited in claim 5 obtained after manipulation of the production host as recited in claim 26.

29. A modified strain as recited in claim 28 that is *Streptomyces lividans*.

30. An *E. coli-Streptomyces* artificial chromosome as recited in claim 3 that carries the erythromycin

gene cluster.

31. An *E. coli-Streptomyces* artificial chromosome as recited in claim 30 that contains the vectors pPAC-S1 or pPAC-S2.

32. An actinomycete production host as recited in claim 2 that contains the *E. coli-Streptomyces* artificial chromosome carrying the rifamycin gene cluster.

33. An actinomycete production host as recited in claim 32 that is *Streptomyces lividans*.

34. A modified strain as recited in claim 5 obtained after manipulation of the production host as recited in claim 32.

35. A modified strain as recited in claim 34 that is *Streptomyces lividans*.

36. An *E. coli-Streptomyces* artificial chromosome as recited in claim 3 that carries the rifamycin gene cluster.

37. An *E. coli-Streptomyces* artificial chromosome as recited in claim 36 that contains the vectors pPAC-S1 or pPAC-S2.

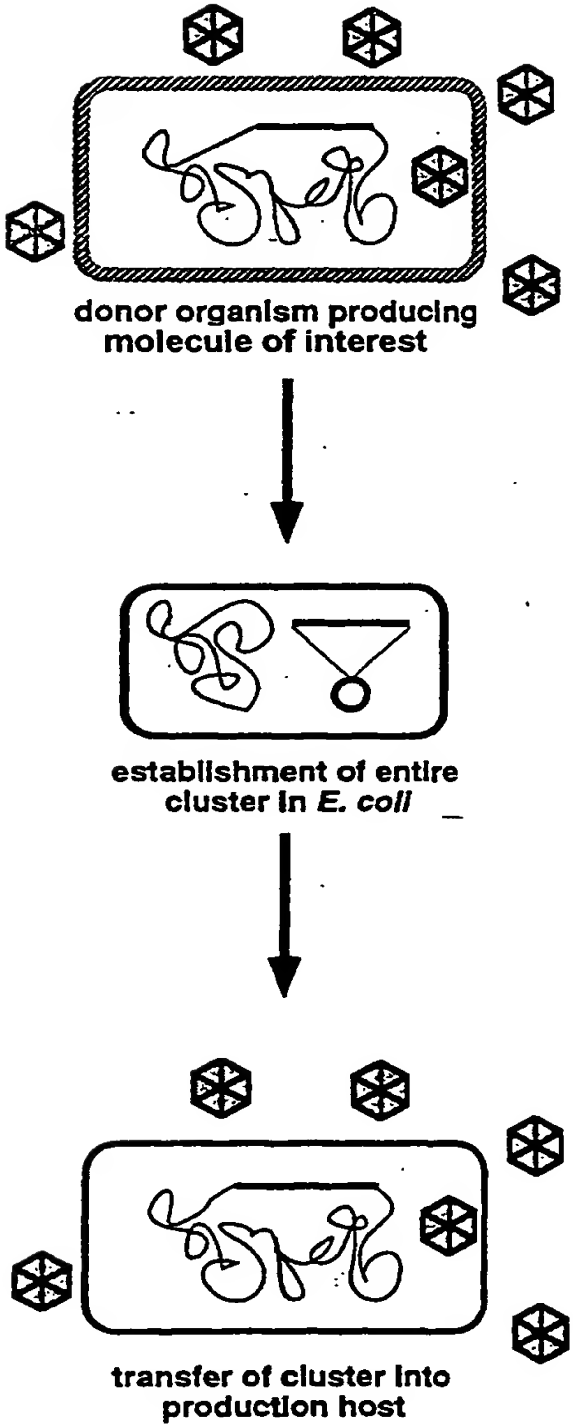

Figure 1

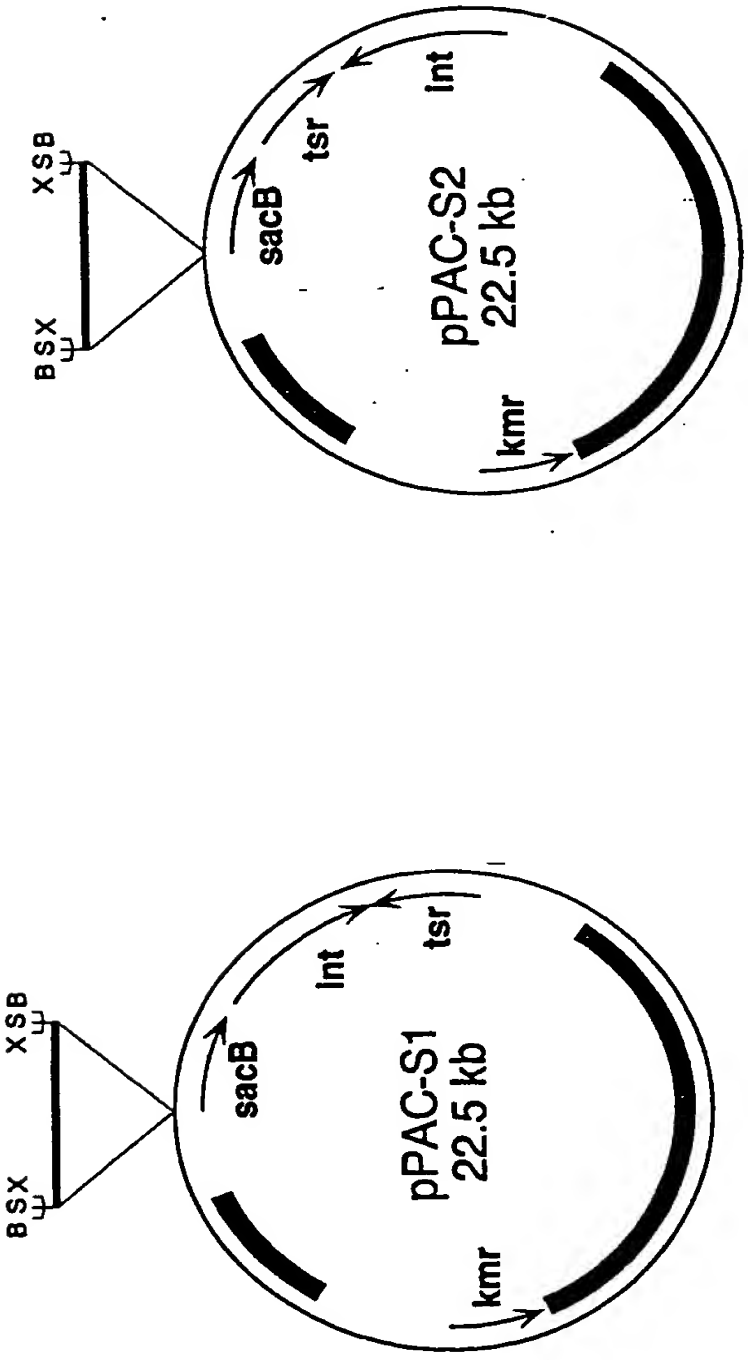
XSB
BSX
sacB
tsr
Int
pPAC-S2
22.5 kb
kmr
XSB
BSX
sacB
Int
tsr
pPAC-S1
22.5 kb
kmr

Figure 2

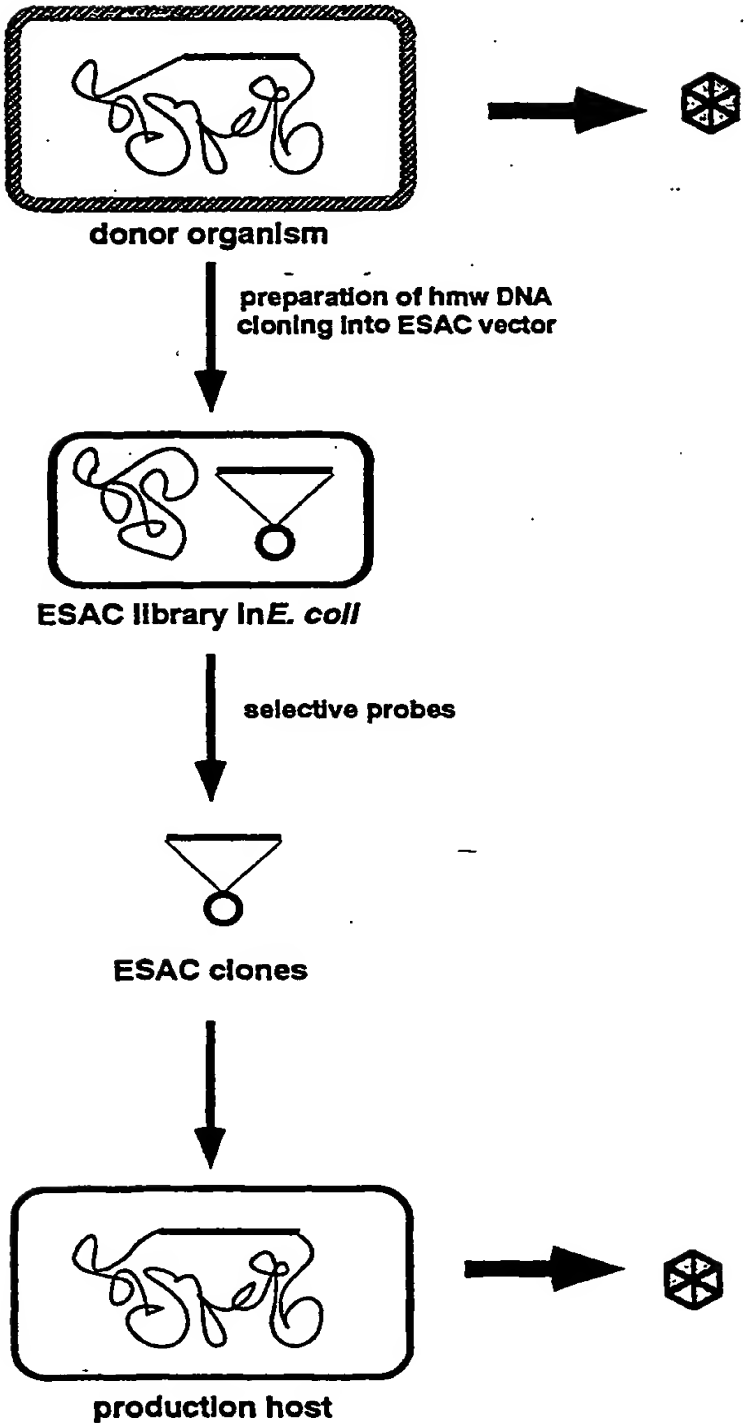
donor organism
preparation of hmw DNA
cloning into ESAC vector
ESAC library in E. coli
selective probes
ESAC clones
production host

Figure 3

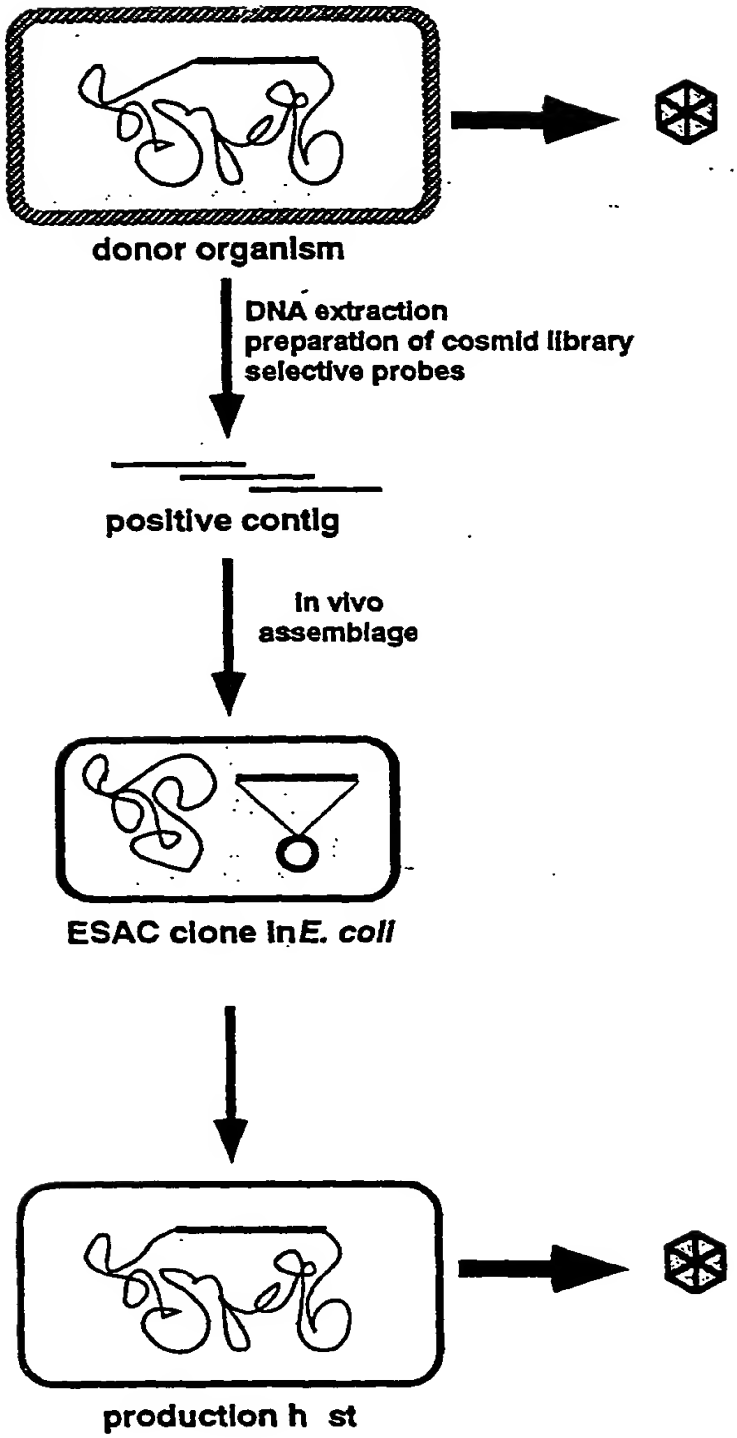
donor organism
DNA extraction
preparation of cosmid library
selective probes
positive contig
In vivo
assemblage
ESAC clone in E. coli
production h st

Figure 4

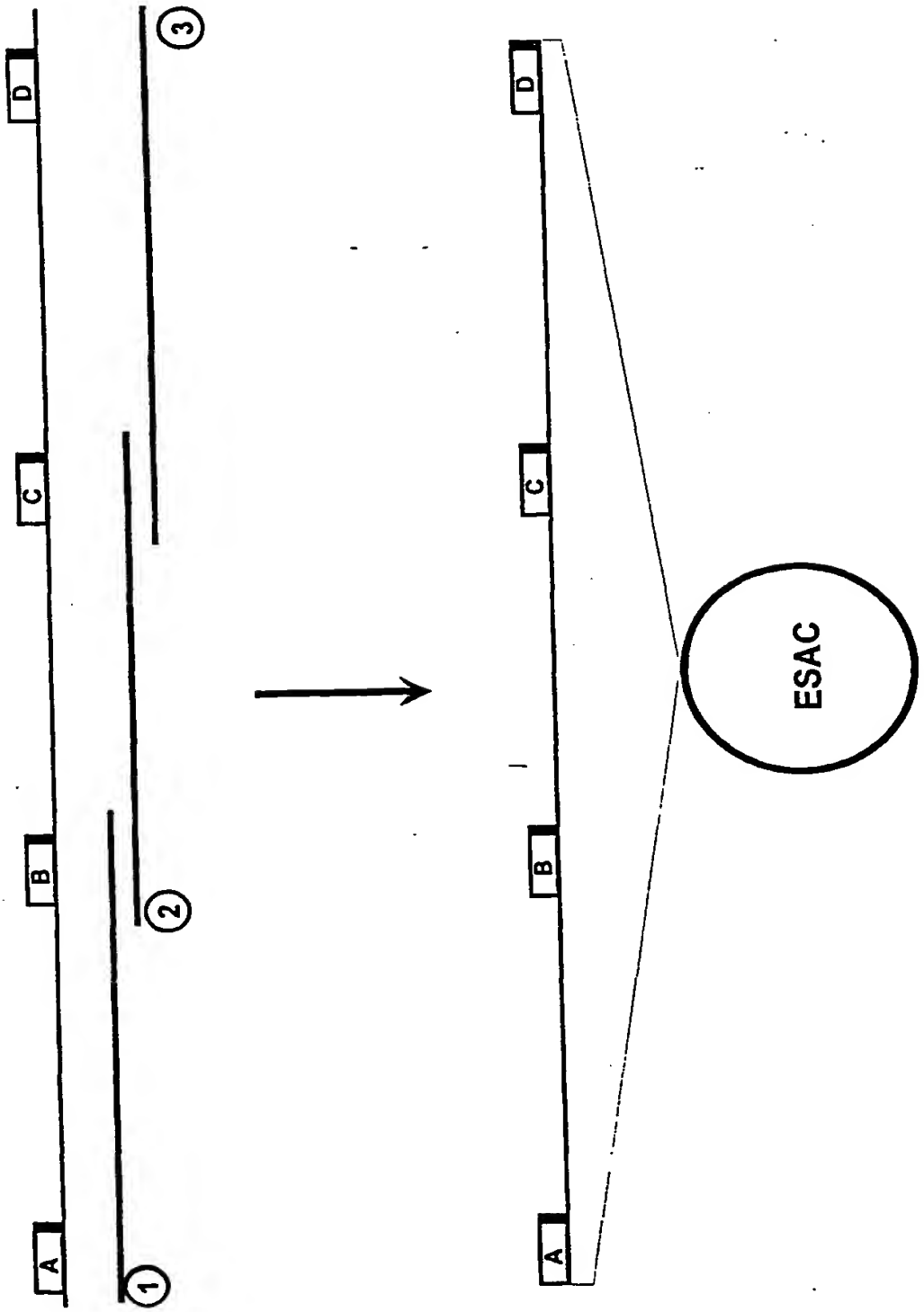

Figure 5

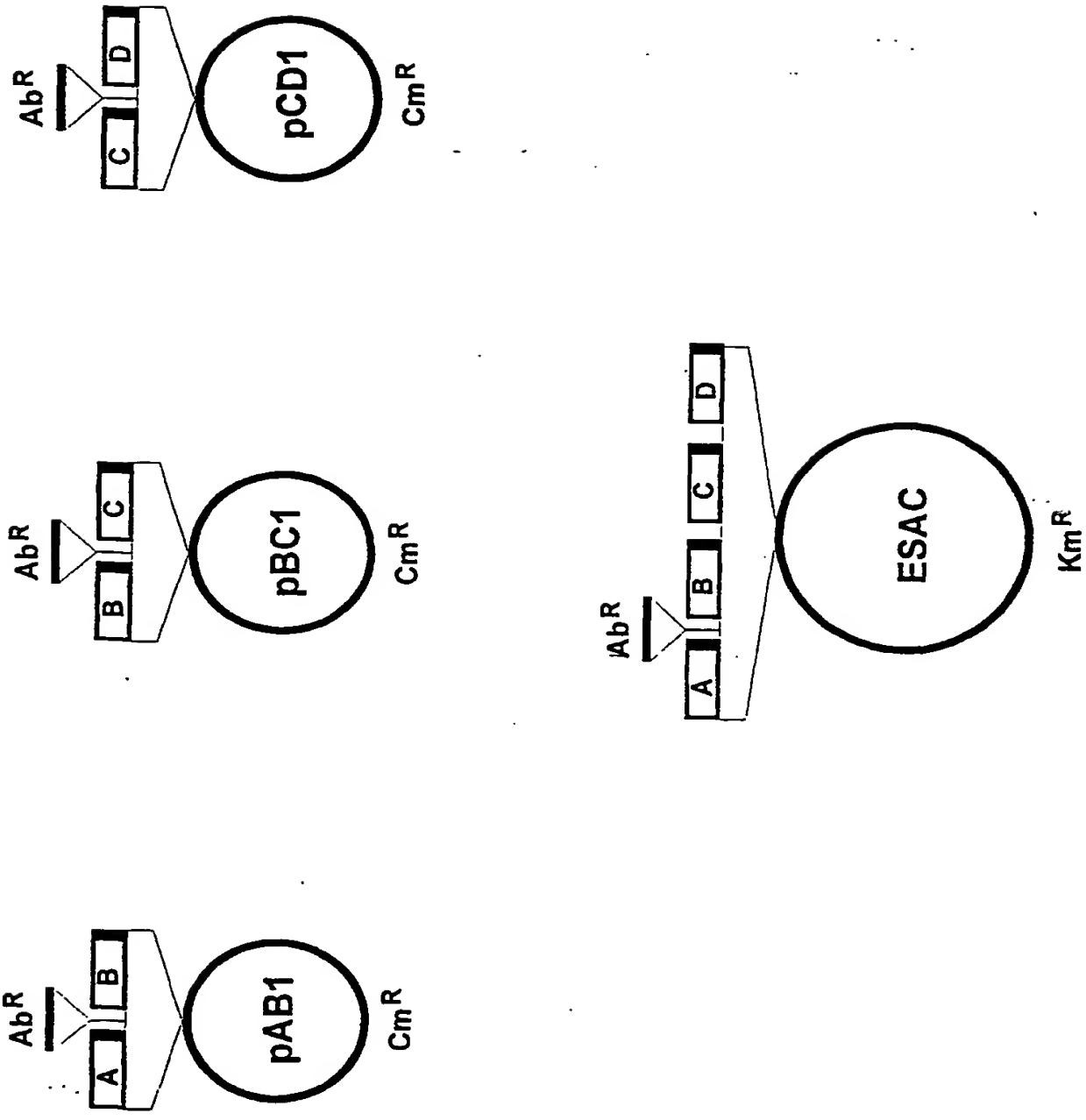
Ab^R
D
C
pCD1
Cm^R
Ab^R
C
B
pBC1
Cm^R
Ab^R
A
B
pAB1
Cm^R
Ab^R
A
B
C
D
ESAC
Km^R

Figure 6

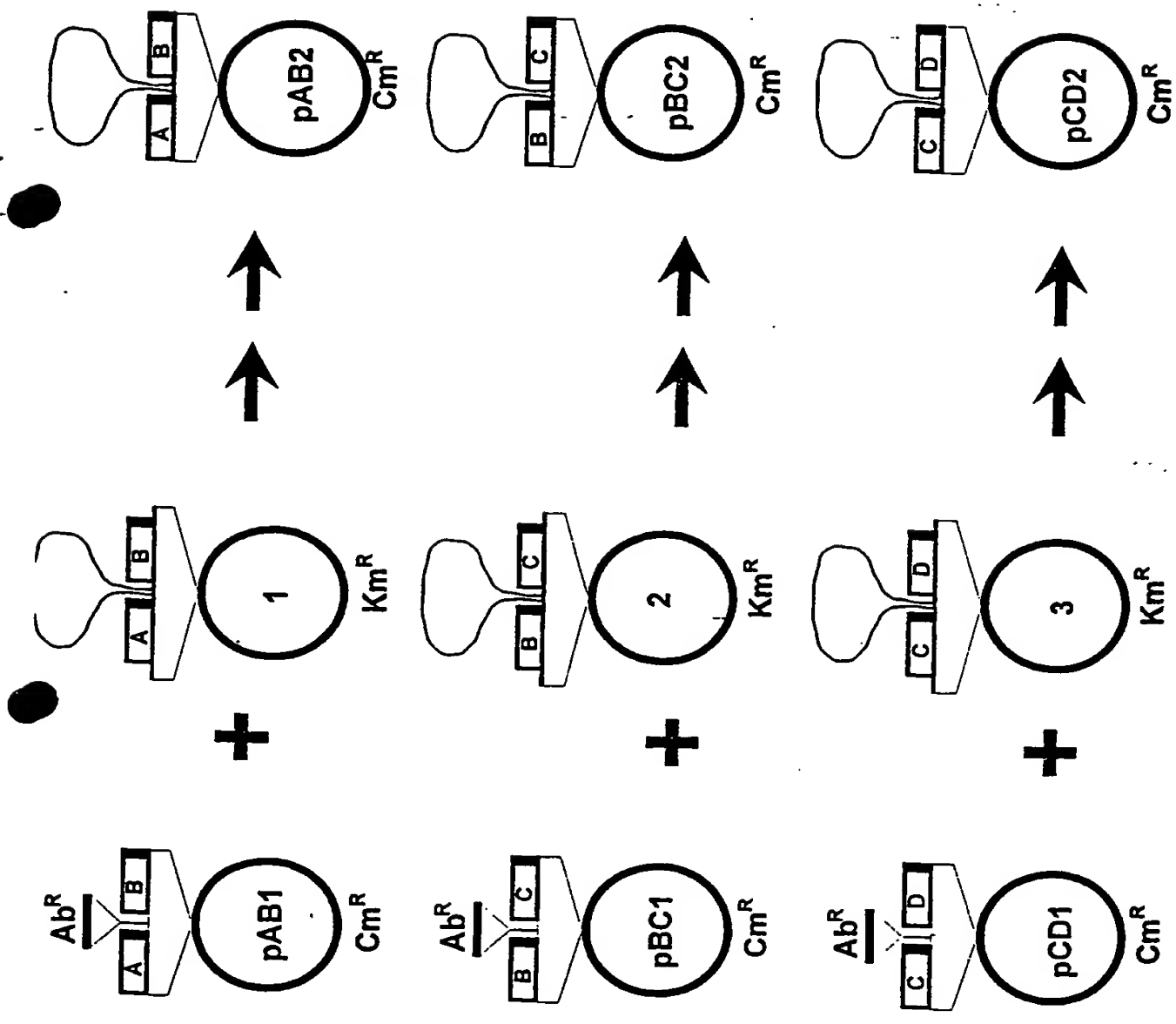

Figure 7

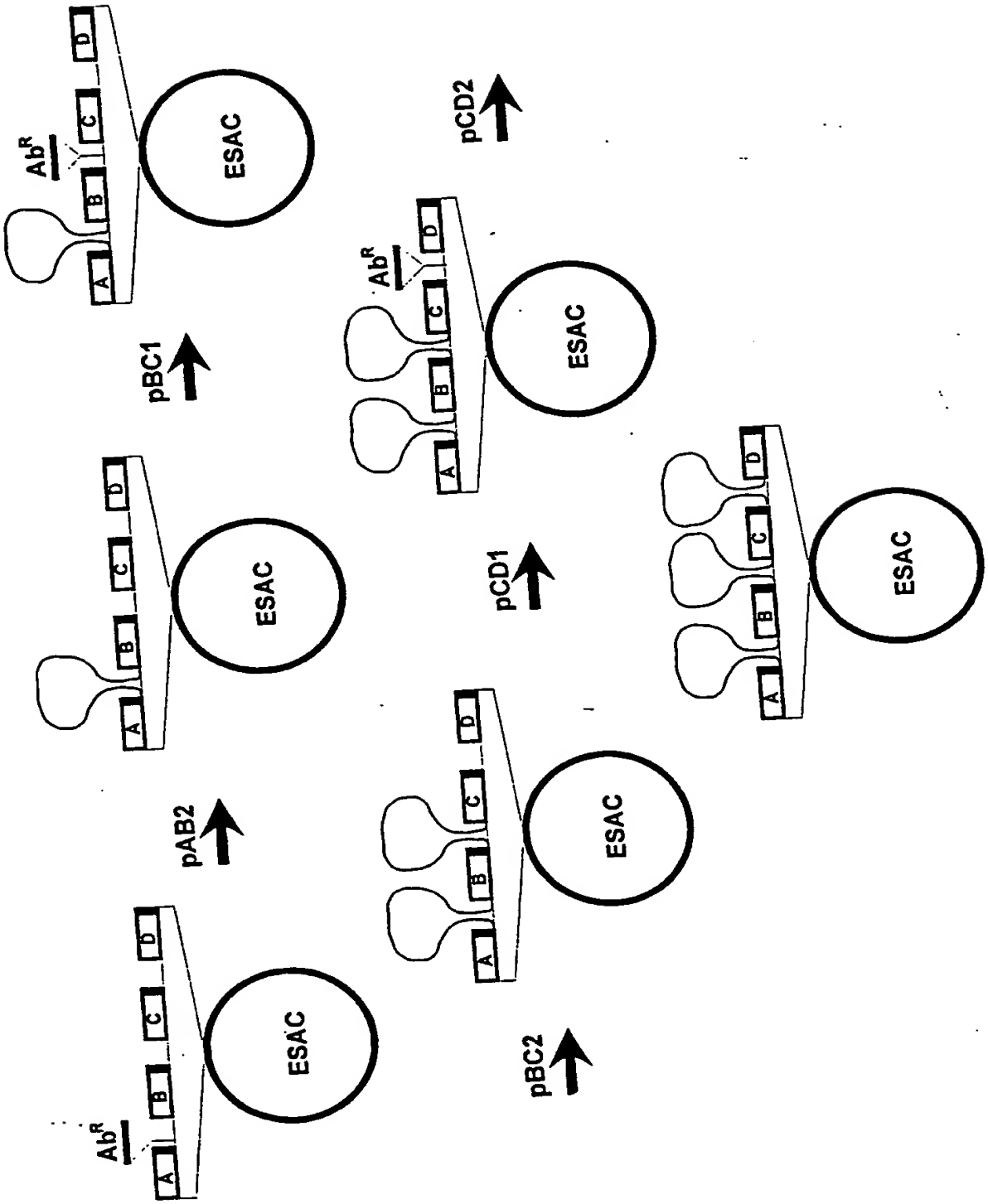

Figure 8

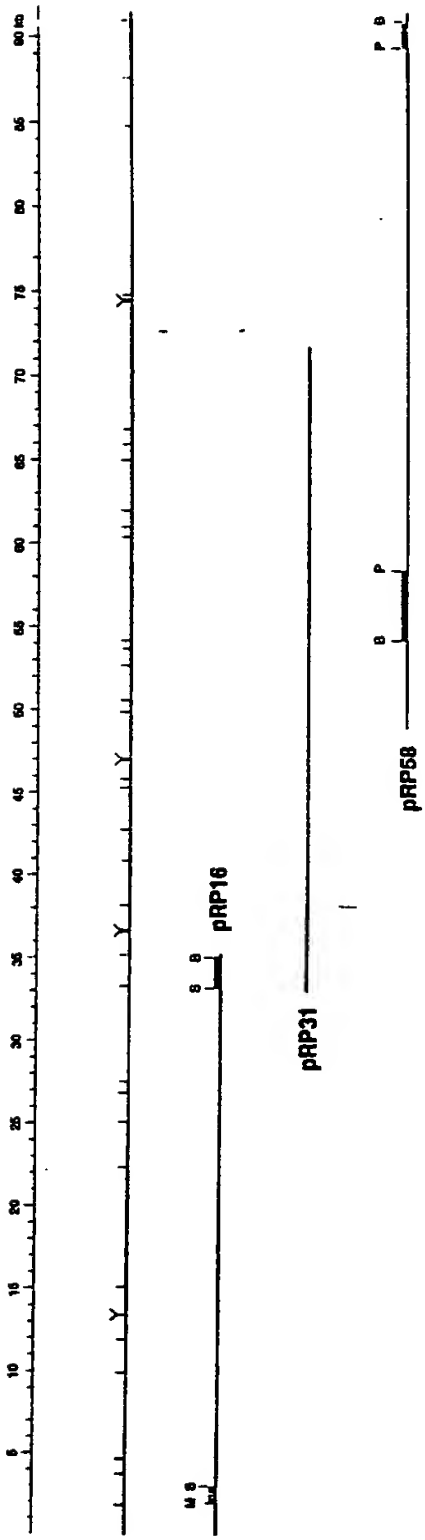

Figure 9

A

```
  1  GGATCCCGAG CACCGACCAG CCGTGGGCGG GGACGAGACA CGGGTCTCCC
 51  GGAGCCTCCC CCGACGACTC CAGCACGGCC AGGCCCGCGG CCTCGACCGG
101  GAAGCGGTAG GGCCTGTCGT CCACGGTTGA GCAGGGTGAG CAGTGCCCGG
151  CCGGGATGGT CCGGGTCAGC CGAGGCCAGC GCGGCGGCCC GGTTGCTCAG
```

B

```
  1  CCGGGAGATC CGCCGACGCC GGCGGCCGTG CACCACGGTC CTCCTGTTCC
 51  GGGCGACGGT GAACGGCAGG CAGGTCCACG GTTCCGATTT CCTGCACTTC
101  GACGACGACG GCCTCATCGG CGAGCTCACC GTCATGGTCC GGCCGCTGTC
151  GGGAGCGAAC GCGCTGGCCG AGGCGATGGG CGCCCAGTTC GAACGGATCC
```

Figure 10

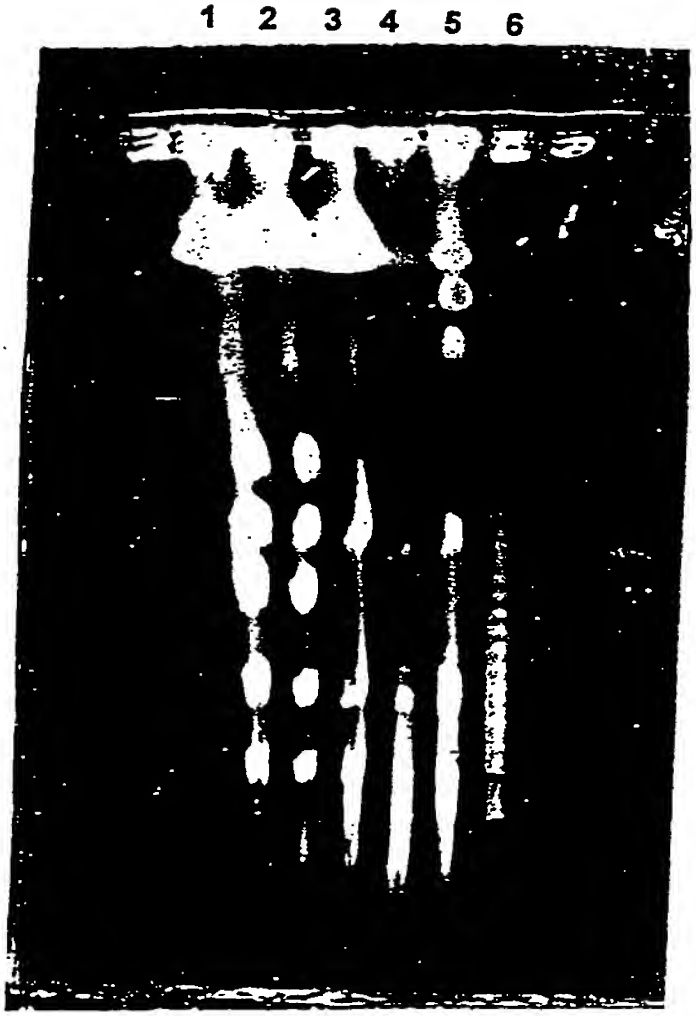

Figure 11

12/12

Figure 12

Abstract

The present invention provides a system for producing and modifying natural products produced by a large group of bacteria for the purpose of drug discovery, development and production. The method of the invention transfers the ability to produce a secondary metabolite from an actinomycete that is the original producer of the natural product, to a different production host that has desirable characteristics. The system involves the construction of a segment of the chromosome of the original producer in an artificial chromosome that can be stably maintained in a suitable production host. The present invention relates to recombinant DNA vectors useful for shuttling the genetic information necessary to synthesize a given natural product between a donor organism and a production host. The methods of the invention are useful in improving the yield, the purification process and for structural modification of a natural product.